



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 117402**

**TO: Manjunath N Rao**  
**Location: REM/3B81/3C70**  
**Art Unit: 1652**  
**Monday, March 29, 2004**

**Case Serial Number: 10/026140**

**From: Noble Jarrell**  
**Location: Biotech-Chem Library**  
**Rem 1B71**  
**Phone: 272-2556**

**Noble.jarrell@uspto.gov**

### **Search Notes**

**This Page Blank (uspto)**

From: Rao, Manjunath N.  
Sent: Monday, March 22, 2004 11:41 AM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search request for 10/026,140

From: Manjunath N. Rao  
Art Unit 1652, Room 3B81  
Mail Box in Room 3C70  
Phone: 272-0939

Date: 3-22-04

Please search the following as soon as possible for application with serial number

**10/026,140**

1. SEQ ID NO: 1 and 3 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2, against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao  
Art Unit 1652, Room 3B81,  
Mail Box in Room 3C70,  
Remsen Building, USPTO

Searcher: Janrell  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 3/29/04  
Searcher Prep/Review: 3  
Clerical: \_\_\_\_\_  
Online time: 60

TYPE OF SEARCH:  
NA Sequences: 2  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: Compu gen  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

RECEIVED  
MAR 22 2004  
STIC

**This Page Blank (uspto)**



400, Dulany St.  
Alexandria, VA.  
Phone: 571-272-0939

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**This Page Blank (uspto)**



# STIC SEARCH RESULTS FEEDBACK FORM

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor  
571-272-2507 Remsen E01 D86

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.





**This Page Blank (uspto)**



```
Db 121 |T|Y|H|M|D|P|O|A|L|H|D|R|Y|G|G|W|L|N|V|E|V|Q|D|F|E|R|Y|A|R|L|C|E|R|G|D|V|Q|W|I|T|I|N|X|W|I|Q|A|I|Y|G| 180
Qy 181 |A|G|S|N|A|P|G|S|S|I|N|K|S|T|E|G|N|T|A|T|E|P|L|A|K|A|Q|I|M|S|H|A|R|A|V|A|Y|S|D|F|R|P|S|Q|G|I|G|I|S|L|N| 240
Db 181 |A|G|S|N|A|P|G|S|S|I|N|K|S|T|E|G|N|T|A|T|E|P|L|A|K|A|Q|I|M|S|H|A|R|A|V|A|Y|S|D|F|R|P|S|Q|G|I|G|I|S|L|N| 240
Qy 241 |G|D|Y|E|P|W|D|S|N|E|P|R|D|K|E|A|R|R|M|E|F|H|I|G|W|A|N|P|I|L|K|O|Y|P|E|S|M|K|O|L|E|R|L|P|A|L|P|A|P|A| 300
Db 241 |G|D|Y|E|P|W|D|S|N|E|P|R|D|K|E|A|R|R|M|E|F|H|I|G|W|A|N|P|I|L|K|O|Y|P|E|S|M|K|O|L|E|R|L|P|A|L|P|A|P|A| 300
Qy 301 |I|N|A|G|E|T|P|Y|G|M|N|Y|T|S|Q|F|A|R|H|D|G|P|V|E|T|D|Y|L|G|A|I|H|E|O|E|N|K|D|G|S|P|V|G|E|S|G|L|A|M|L|R|S|C| 360
Db 301 |I|N|A|G|E|T|P|Y|G|M|N|Y|T|S|Q|F|A|R|H|D|G|P|V|E|T|D|Y|L|G|A|I|H|E|O|E|N|K|D|G|S|P|V|G|E|S|G|L|A|M|L|R|S|C| 360
Qy 361 |P|M|F|R|K|H|A|R|V|Y|G|L|G|K|P|I|Y|I|T|E|N|G|C|P|C|G|E|E|N|M|T|C|E|A|V|N|D|P|R|I|R|Y|F|D|S|H|L|D|S|I|S|K|A|I| 420
Db 361 |P|M|F|R|K|H|A|R|V|Y|G|L|G|K|P|I|Y|I|T|E|N|G|C|P|C|G|E|E|N|M|T|C|E|A|V|N|D|P|R|I|R|Y|F|D|S|H|L|D|S|I|S|K|A|I| 420
Qy 421 |T|O|D|G|V|V|V|K|G|Y|F|A|M|A|L|L|D|N|L|E|W|S|D|G|Y|G|P|R|G|V|T|F|T|D|Y|T|L|K|R|T|P|K|S|A|L|V|L|K|D|M|F|A|R|Q|R|V| 480
Db 421 |T|O|D|G|V|V|V|K|G|Y|F|A|M|A|L|L|D|N|L|E|W|S|D|G|Y|G|P|R|G|V|T|F|T|D|Y|T|L|K|R|T|P|K|S|A|L|V|L|K|D|M|F|A|R|Q|R|V| 480
Qy 481 |K|V|A|A| 484
Db 481 |K|V|A|A| 484
```

## RESULT 2

```
ID 093784 PRELIMINARY; PRT; 476 AA.
AC 093784;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Beta-glucosidase.
GN BGL4.
OS Humicola grisea var. thermoides.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP SEQUENCE FROM N.A.
RA Takashima S., Nakamura A., Hidaka M., Masaki H., Uozumi T.,
RT "Molecular cloning and expression of the novel fungal beta-glucosidase
RL genes from Humicola grisea and Trichoderma reesei."
J. Biochem. 0:0-0(1999).
DR EMBL; AB003109; BAAT4958.1; -.
DR HSSP; P22073; ITR1.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
SQ SEQUENCE 476 AA; 54062 MW; DA17ADF5CBB9935 CRC64;
```

Query Match 49.3%; Score 1303.5; DB 3; Length 476;

Best Local Similarity 51.7%; Pred. No. 2.5e-96;

Matches 244; Conservative 79; Mismatches 144; Indels 5; Gaps 5;

```
Qy 5 |L|A|L|P|N|D|E|W|G|P|A|T|A|Y|O|I|E|G|A|V|E|G|G|S|I|N|D|Y|C|H|L|E|R|S|R|N|G|A|N|G|D|V|A|C|H|Y|R|Y|D|E| 64
Db 1 |M|S|L|P|D|K|K|G|F|A|T|A|Y|O|I|E|G|S|V|E|D|G|G|S|I|W|D|T|P|C|A|I|P|G|K|I|A|D|G|S|G|A|V|A|C|S|Y|R|K|T|Y|E| 60
Qy 65 |D|F|L|L|T|K|Y|G|A|K|A|V|F|S|L|S|W|R|I|I|P|L|G|R|L|D|P|V|N|E|G|I|E|F|S|K|L|D|A|L|R|R|G|I|P|W|T|L|V|H| 124
Db 61 |D|I|A|L|L|K|E|L|G|A|N|S|Y|R|F|S|I|S|W|R|I|I|P|L|G|R|N|D|P|I|N|K|G|I|D|H|V|K|F|V|D|L|I|E|G|I|P|I|T|L|R|H| 120
Qy 125 |W|D|L|P|A|L|H|D|Y|G|G|W|L|N|V|E|V|Q|D|F|E|R|Y|A|R|L|C|E|R|F|D|R|V|Q|W|I|T|I|N|X|W|I|Q|A|I|Y|G|A|T|S| 184
Db 121 |W|D|L|P|A|L|H|D|Y|G|G|W|L|N|V|E|V|Q|D|F|E|R|Y|A|R|L|C|E|R|F|D|R|V|Q|W|I|T|I|N|X|W|I|Q|A|I|Y|G|A|T|S| 179
```

```
Qy 185 |N|A|P|R|S|S|I|N|K|S|T|E|G|N|T|A|T|E|P|L|A|K|A|Q|I|M|S|H|A|R|A|V|A|Y|S|D|F|R|P|S|Q|G|I|G|I|S|L|N|D|Y| 244
Db 180 |P|A|P|H|T|D|R|K|S|P|V|G|S|A|R|E|P|I|V|G|H|I|L|A|H|A|R|A|V|A|Y|E|D|E|K|P|O|G|G|E|I|G|I|T|L|N|D|A|T| 239
Qy 245 |E|P|W|D|S|N|E|P|R|D|K|E|A|R|R|M|E|F|H|I|G|W|A|N|P|I|L|K|O|Y|P|E|S|M|K|O|L|E|R|L|P|A|L|P|A|P|A|I|N|A| 304
Db 240 |L|P|W|D|P|A|I|E|A|C|D|R|I|E|R|A|I|S|F|A|D|P|I|F|G|K|-Y|P|S|M|K|O|L|G|D|R|L|E|P|T|E|R|V|A|L|V|K|- 297
Qy 305 |G|E|T|D|F|Y|G|M|N|Y|T|S|Q|F|A|R|H|D|G|P|V|E|T|D|Y|L|G|A|I|H|E|O|E|N|K|D|G|S|P|V|G|E|S|G|L|A|M|L|R|S|C|P|D|M|F| 364
Db 298 |G|S|N|D|F|Y|G|M|N|Y|T|A|N|Y|I|K|H|K|T|G|V|P|E|D|F|G|N|L|E|T|L|F|N|K|Y|G|D|C|I|G|P|E|T|G|S|F|W|R|P|H|A|Q|F| 357
Qy 365 |R|K|H|A|R|V|Y|G|L|G|K|P|-Y|I|T|E|N|G|C|P|C|G|E|E|N|M|T|C|E|A|V|N|D|P|R|I|R|Y|F|D|S|H|L|D|S|I|S|K|A|I|T|O|D| 423
Db 358 |R|D|L|N|M|S|K|R|Y|G|P|K|I|Y|V|E|N|G|S|L|K|E|N|D|M|P|E|Q|V|L|E|D|F|R|Y|K|Y|F|N|D|Y|R|A|A|A|V|A|E|D| 417
Qy 424 |G|V|V|V|K|G|Y|F|A|M|A|L|L|D|N|L|E|W|S|D|G|Y|G|P|R|G|V|T|F|T|D|Y|T|L|K|R|T|P|K|S|A|L|V|L|K|D|M|F|A|R|Q|R|V| 474
Db 418 |G|N|V|R|G|I|A|S|L|D|N|F|E|W|A|G|E|T|R|G|V|T|Y|V|Y|A|N|Q|K|Y|P|K|S|A|S|L|A|P|L|F| 469
```

## RESULT 3

```
ID 08X214 PRELIMINARY; PRT; 489 AA.
AC 08X214;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Beta-glucosidase 1 (EC 3.2.1.21).
GN BGL.
OS Talaromyces emersonii.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OX Eurotiaceae; Trichocomaceae; Talaromyces.
OX NCBI_TaxID=68825;
RN [1]
RP SEQUENCE FROM N.A.
RA Murray P.G., Collins C.M., Tuohy M.G.;
RT "Molecular cloning and expression analysis of beta glucosidase from
RL Talaromyces emersonii."
Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF439322; AAL34084.2; -.
DR EMBL; AY081764; AAL89551.2; -.
DR GO; GO:0004422; P:beta-glucosidase activity; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolase; Glycosidase.
SQ SEQUENCE 489 AA; 55810 MW; CF438E4ED72D52BA CRC64;
```

Query Match 49.1%; Score 1298; DB 3; Length 489;

Best Local Similarity 52.2%; Pred. No. 7.3e-96;

Matches 246; Conservative 70; Mismatches 149; Indels 6; Gaps 6;

```
Qy 7 |L|N|D|E|W|G|P|A|T|A|Y|O|I|E|G|A|V|E|G|G|S|I|N|D|Y|C|H|L|E|R|S|R|N|G|A|N|G|D|V|A|C|H|Y|R|Y|D|E|F| 66
Db 15 |L|P|D|L|W|G|F|A|T|A|Y|O|I|E|G|A|F|D|E|G|L|P|S|I|W|D|T|F|S|T|P|K|E|V|E|D|G|N|G|D|V|A|C|S|Y|R|H|T|G|D|I| 74
Qy 67 |D|L|T|Y|G|A|K|A|V|F|S|L|S|W|R|I|I|P|L|G|R|L|D|P|V|N|E|G|I|E|F|S|K|L|D|A|L|R|R|G|I|P|W|T|L|V|H|D| 126
Db 75 |E|L|T|K|Y|G|A|K|I|R|F|S|L|S|W|R|I|I|P|L|G|R|N|D|P|I|N|K|G|I|D|H|V|K|F|V|D|L|I|E|G|I|P|I|T|L|R|H|D| 134
Qy 127 |L|P|A|L|H|D|Y|G|G|W|L|N|V|E|V|Q|D|F|E|R|Y|A|R|L|C|E|R|F|D|R|V|Q|W|I|T|I|N|X|W|I|Q|A|I|Y|G|A|T|S|N|A| 186
Db 135 |L|P|E|L|M|K|Y|G|G|W|L|N|V|E|V|Q|D|F|E|R|Y|A|R|L|C|E|R|F|D|R|V|Q|W|I|T|I|N|X|W|I|Q|A|I|Y|G|A|T|S|N|A| 194
Qy 187 |P|R|S|S|I|N|K|S|T|E|G|N|T|A|T|E|P|L|A|K|A|Q|I|M|S|H|A|R|A|V|A|Y|S|D|F|R|P|S|Q|G|I|G|I|S|L|N|D|Y|E|P| 246
Db 195 |P|R|T|S|D|R|K|S|P|E|G|D|T|R|E|P|I|V|G|H|I|L|A|H|A|R|A|V|A|Y|E|D|E|K|P|O|G|G|E|I|G|I|T|L|N|D|W|A|E|P| 254
```

Oy		247	WDSNEPRCKEAAERMEEHGIFANPIPLKDDYPSRMKKQGERPALTPADFAILNMG	306
Dd		255	WDPENPADVEACXIERIEPAISMFADPIHNG-YIPSMVKQLDRLPKPTPEIAFVH-GS	312
Oy		307	TDFYGMNYTTSQCFARHLHDG-DVPETDYLGAIIHEOENVDGSPVGESGLAWLRSCPDMPR	365
Dd		313	NDFPGMNHHCENYIRNRTEGDDE-DIANGNDLIMEDKNGNINIGETOCCEWLRRPFLGR	371
Oy		366	KHLARVGLYKRP-IYTENGPCPCGEENMTCEAVNDPFRIYRDSHLSISKAITODG	424
Dd		372	KLLKMLADRNVNNPKYITVNTGTSVKGESDKPLEEVLANDEFVQYRYDYIAGAVDAVDG	431
Oy		425	VVVGYPMALLDNLEMSDGCPRGCVFTDTYT-LKXTPKKSALVLD MF	474
Dd		432	VVVKAYMAWSLLDNFESEEGRSRGVTYVDIKNQOKRI PKXSALVIGELF	482
 RESULT 4 093785 PRELIMINARY; PRT; 466 AA.				
ID	093785			
AC	093785;			
DT	01-MAY-1999 (TREMBLrel). 10, Created)			
DT	01-MAY-1999 (TREMBLrel). 10, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel). 24, Last annotation update)			
DE	Bete-glucosidase.			
GN	BGL2.			
OS	Trichoderma reesei (Hypocrea jecorina).			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
OX	NCBI_TaxID=51453;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=QM9414;			
RA	Takashima S.; Nakamura A., Hidaka M., Masaki H., Uozumi T.;			
RT	"Molecular cloning and expression of the novel fungal beta-glucosidase			
RL	genes from Humicola grisea and Trichoderma reesei.";			
RL	J. Biochem. 0:0-0(1999).			
DR	EMBL; AB003110; BAAV4959.1; --			
DR	HSSP; P22073; ITR1.			
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro: IPRO01360; Glyco_hydro_1.			
DR	Pfam: PF00232; Glyco_hydro_1; 1.			
DR	PRINTS; PR00131; GLHYDRLASE1.			
DR	Prodrom; PD000650; Glyco_hydro_1; 1.			
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.			
SO	SEQUENCE 466 AA; 52240 MW; 73CAB66FCOA265B7 CRC64;			
 Query Match 48.0%; Score 1270; DB 3; Length 466; Best Local Similarity 52.7%; Pred. No. 1.2e-93; Matches 248; Conservative 65; Mis-matches 142; Indels 16; Gaps 7				
Oy		7	LPNDEMFPATPAVOIEGAVKEGGKGPSIWDTYCHLEPERBTNGANGDVACHYHRDEDF	66
Dd		2	LPKDQWFAPATAVOIEGVADDDGSGPSIWDFFCAQPGKIADSGSGVTA CDSYNTRAEDI	61
Oy		67	DLLTKYGAAYAFSLSSWRSIIPLGRLDPVNEEGIEFYSKLIDALLRGIPTWVLTYHW	126
Dd		62	ALLKSLGAKSRFSISWRSRIIEPGRGADVANNAGDIHVYKFVDDL DLAGITPTFILFHMD	121
Oy		127	LPLALHRIYGMNLNVEVOLDBERYARLCBFERFGORVONMTINKPWIALIYGATGSNA	186
Dd		122	LPEGIHQHYGGILLNTEPFLDENYARVWF-FALPKVRNWITFNBP LCSAIGYSGSTPA	180
Oy		187	PGRSSINHGSTGNATPEWLA GKAOIWSHARA VAVYSMDPP-SQGQIGISLNGDYXE	245
Dd		181	PGROS-----TSEPWTYGHNIILVAHGRAWAYRDPDFPASDDGIGIVLNGDFY	230
Oy		246	PWDSNEPRDKEAERREMEFHIGFANPIPLKDDYPESMKQUGERLPALTTPADFALNAG	305
Dd		231	PWDADPADPADKEAERLERLEFTMWADPYTL-GDIYASNMKQIGDRPLPTTPBERALVH-G	288
Oy		306	ETDFYGMNYTTSQCFARHLHDGPVETDYLGAIIHEOENVDGSPVGESGLAWLRSCPDMPR	365

Dd		269	SNDYVGNMHTSNIRIRSSP	ASADTVGVAVDVLFTNKQNCNGPEFGRPSMLRCAAGFR	348
Oy		366	KHLARVYLIG-KPIYTENGCCPCGGEANMTCEBAVNDPRIRKYPDSHLSISAKITQDG		424
Dd		349	DFLWMIKRKYGPPIYVTENGTSTIKGESDLPKKEKLIEDDERVKXNYNEIYAMTAVAEIDG		408
Oy		425	VVVVGAYAMLLDNLEMSDCYGPRFGTPTFDYTT-LKRTFKSALVLCXMF		474
Dd		409	VNVKGYPAMSIMDNFEWADGYVTRFGVTVVDYENGQRFPKSKSLKPLF		459

  

RESULT 5

Q7ZTX1	PRELIMINARY;	PRT:	450 AA.
ID Q7ZTX1			
AC Q7ZTX1;			
DT 01-OCT-2003 (TREMBLrel. 25, Created)			
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)			
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE Beta-glucosidase 2.			
GN BGL2.			
OS Trichoderma viride.			
OC Eukaryota; Fungi; Ascomycetes; Pezizomycotina; Sordariomycetes;			
OC Hypococcmyceteidae; Hypococcales; mitosporic Hyphocreales; Trichoderma.			
NCL TaxID=5547;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=AS 3.3711;			
RA Liu B.D., Yang O., Zhou Q.;			
RT "Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from			
RL Trichoderma viride AS 3.3711."			
SR Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AY343988; AAQ21384.1; -			
SO SEQUENCE 450 AA; 50020 MW; 9C9A5A1EBCF28B82 CRC64;			

  

Query Match	44.3%	Score 1172;	DB 3;	Length 450;
Best Local Similarity	49.9%;	Pred. No. 8.9e-86;		
Matches 236;	Conservative 65;	Mismatches 132;	Indels 40;	Gaps 10;

  

Oy	7	LPNDFEGFATAVAQIEGA VKEGGRGPSINDTYCHLESRTNGANGVACDHVHR YDEDF	66
Dd	2	LKPQFMGFANAQAQIGADV DQGRGSIMDTFCAPQGKIADSGSVTA CDSYNRTAEDI	61
Oy	67	DLTKYGA KAYRPSLSWSRIIP LCGRLDPVNESGIESFYSLIDALLRGITTPWTLVYMD	126
Dd	62	ALLTSIAKSYRFISIRSRIIP EGGRSDAVNAQAGIDHYVVFVNDLLDAGITPPITLFHMD	121
Oy	127	LPOALHRYGGMILVEEVOLDFEEYARLCFERPEDRVOMNTINXPWIOALYGA TGSNA	186
Dd	122	LPEGHORYGGLLRTEPFLDFENYA VMF-RALPKRMNTTFNEPLCSAPIGSGSIFA	180
Oy	187	PGRSINKHSTEGMTATEPMLAGAKQIMSHARA VA VSRDPRP-SQKGOIGISLNGDYE	245
Dd	181	PGRS-----TSEPMTVGHNIILVAGRANKAYRADPFKPA SGQGIIVLNGDFTY	230
Oy	246	PWDSNEPRDKAERMEFHIGWFANDIP LKKOY PESHKIQLGRLPALTPADFA IINAG	305
Dd	231	PMDAADPADKXAARRRLEPFTAWPADPIYL-GDYPASMRKQLGRLTPTFPPEERALVH-G	288
Oy	306	ETDYYGMNXYTSGAFARILDGVPTDV LGAIHENQEKDQSPVEBSGGLMLRSCPMFR	365
Dd	289	SNDYVGNMHTSNIRIRSSPASTT-----PSATTCSSPTSRAVASAPRRSP----	337
Oy	366	KHLARV-----YGLYGRKPIYTTENGCCPCGGEANMTCEBAVNDPRIRKYPDSHLD	414
Dd	338	--LAPVRPAATSMWGSA RGTY-PPIYVTENGTSTIKGESDLPKKEKLIEDDPRVKXNYNYIR	394
Oy	415	SISAKITQDGVVVKGYFAMALLDNLEMSDCYGPRFGTPTFDYTT-LKRTPKS	466
Dd	395	AMTAVAEIDGVNVKGYFAMSIMDNFEWADGYVTRFGVTVVDYENGQRFSPPRA	447

  

RESULT 6

Q9SP9 PRELIMINARY; PRT; 540 AA.  
 ID Q9SP9  
 AC Q9SP9  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 OS Raucolfia serpentina (Serpentine) (EC 3.2.1.125).  
 OS Raucolfia serpentina (Serpentine) (Devilpepper).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC lamids; Gentianales; Apocynaceae; Rauvolfiaceae; Vincet; Rauvolfia.  
 OC NCBI\_TaxID=4060;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BENTH. ex KURZ;  
 RA Warzcha H., Gerasimenko I., Kutchan T.M., Stoeckigt J.;  
 RT "Cloning, bacterial expression, and properties of raucolficine  
 RT glucosidase, an enzyme of indole alkaloid biosynthesis in Rauvolfia  
 RT serpentina";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF149311; AAF03675.1; -.  
 DR HSSP; P26205; 1CBG.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001360; Glyco\_hydro\_1.  
 DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS; PR00131; GLHYDRLASE1.  
 DR PRODOM; PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 DR KEGG; K00000; Glycosylase; Hydrolase.  
 SQ SEQUENCE 540 AA; 60934 MW; 543A24654A4D5E7D CRC64;

Query Match 41.1%; Score 1087; DB 10; Length 540;  
 Best Local Similarity 45.3%; Pred. No. 7.9e-79;  
 Matches 226; Conservative 69; Mismatches 162; Indels 42; Gaps 14;

8 PNDPFGMTATAAYOIEGAVKEGGRPSIMPTCHLEPSR-INGANGDVACHYHYRDEDF 66  
 23 PADFMGSSAYOIEGARGDGRPSIMPTCHLEPSR-INGANGDVACHYHYRDEDF 82  
 67 DLTLYGAKAYRFSLSMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWVTLVHW 125  
 83 NILKMGADAVRFSISMSRIIP-PGRLSGVNGKGINVNNLLDGLANGIKPVTILFW 141  
 126 DLPOALHRYGGMVNEEYQDLDFERYALCFERFGDRVQNMVITINXPMIOAYGATGSN 185  
 142 DVPQALDEYGGFUSPRIVD-DFCEYALCFERFGDRVQNMVITINXPMIOAYGATGSN 200  
 186 AGRSSIN-CHST-----EGNTATEPMLAGKAQIMSHARAVAVSRD 226  
 201 AGRGRSPRPHVNHPTVQHRCTVAPOCICSTGNPTEFPVTHHLLAHAAVELYKXK 260  
 227 FRPBGQGTIGSLNGDYEPWDSNEPRDKEAERMEFHIGWNPFLKDYEPESMKQ 286  
 261 FQROEGGIGISHYTOVMEPWDSNENASIVAEARALDFMLGFWEP-ITSGDYKSKKF 319  
 287 LGERLPALTPDFAIINAGETDFGMVYYSQFARHLDGVPEDIDYGA-----IH- 337  
 320 VGSLLPFRSPQSKMLK-GSYDFGLNVTYASVYTN-----ASTNSSGSSNNFSYNTDIH 373  
 338 EHQENKDSPVGEESGLAWLNSCPDMFRKLARVYGLYGRP-IYITENGCPCEGEENMTG 396  
 374 TYETDRNGVPIPGSGSDMLIYPEGIRKLIVYKTYVNPFLIYITENGVDVKNNTL 433  
 397 BEAVNDPFRIRYPSHLSISKATITOGGVYVYGFAMALLDNLEMSDGYGPRFGVFTDY 456  
 434 SEAKKMDRLKTYLQDHIENVRQAM-NDGVVYKGYFAWSLIDNFEWEGGVVRFGLIHDY 492  
 457 T-TLKRTPKKSALVTKDMF 474  
 493 NDNFARVPRKDSAVWLMSF 511

RESULT 7  
 ID Q8L7J2  
 AC Q8L7J2  
 DT 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
 OS Beta-glucosidase.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Euharoidae; Oryzae; Oryza.  
 OC NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Yukiikari;  
 RA Akiyama T., Ogasiri R., Ketudat-Cairns J.R., Swast J., Been A.;  
 RT "Predominant expression of beta-glucosidase in germinating rice  
 RT seeds";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY129284; AA001354.1; -.  
 DR Gramene; Q8L7J2; -.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001360; Glyco\_hydro\_1.  
 DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS; PR00131; GLHYDRLASE1.  
 DR PRODOM; PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 DR KEGG; K00000; Glycosylase; Hydrolase.  
 SQ SEQUENCE 521 AA; 58539 MW; 0554435869F3BD21 CRC64;

Query Match 39.7%; Score 1049; DB 10; Length 521;  
 Best Local Similarity 45.1%; Pred. No. 8.6e-76;  
 Matches 214; Conservative 79; Mismatches 157; Indels 24; Gaps 10;

6 ALPNDPEMGATAYOIEGAVKEGGRPSIMPTCHLEPSR-INGANGDVACHYHYRDEDF 65  
 49 SPEGCVFVFGASAAAYEGAVKEDGRTIMPTATPFTKIDFSNADVAVDQHYRFEED 108  
 66 FDLITLYGAKAYRFSLSMSRIIPLGRLD-PVNEBGEFFYSKIDALLRGITPWVTLVHW 125  
 109 IQLMADMGADAVRFSISMSRIIP-PGRLSGVNGKGINVNNLLDGLANGIKPVTILFW 166  
 126 DLPOALHRYGGMVNEEYQDLDFERYALCFERFGDRVQNMVITINXPMIOAYGATGSN 185  
 167 DLPQALDEYGGFUSPRIVD-DFEAVALCFERFGDRVQNMVITINXPMIOAYGATGSN 225  
 186 AGRSSIN-CHST-----EGNTATEPMLAGKAQIMSHARAVAVSRD 243  
 226 AGRGRSPRPHVNHPTVQHRCTVAPOCICSTGNPTEFPVTHHLLAHAAVELYKXK 285  
 244 YEPWDSNEPRDKEAERMEFHIGWNPFLKDYEPESMKQGERLPALTPDFAIIN 303  
 286 FRPBGQGTIGSLNGDYEPWDSNEPRDKEAERMEFHIGWNPFLKDYEPESMKQ 342  
 304 ASETDPYGNVYTSQFARHLDGVPEDIDYGA-----EKKDSPVGEESGL 354  
 343 KQALPFGVGNHYTYTTH-----NNTNIGTLNNTLADGTGVSLEPKNGKPIGDRANS 397  
 355 AWRGCPDMFRKLARVYGLYGRP-IYITENGCPCEGEENMTCEAVNDPFRIRYPSH 413  
 398 IWLIVPRKMSLWYVERVNSPVYITENGMDNSNPFISIDALDSKRIKHYNDL 457  
 414 DSISKATITOGGVYVYGFAMALLDNLEMSDGYGPRFGVFTDY-TLKRTPKKS 466  
 458 TWLAASIKEDGDVAGFYAWSLIDNFEWEGGVVRFGLIHDY 511

RESULT 8  
 ID Q9P456  
 AC Q9P456  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)



DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)  
 DE Beta-glucosidase precursor.  
 GN BGL.  
 OS Aspergillus niger.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 NCBI\_TaxID=5061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AMS11;  
 RA Al Y.-C., Meng F.-M.;  
 RT "Molecular cloning and sequencing of a beta-glucosidase gene from  
 RT Aspergillus niger AMS11."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF269911; AAF74209.2; -.  
 DR HSP; P26205; ICBG.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001360; Glyco\_hydro\_1.  
 DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS; PR00131; GHYDRLASE1.  
 DR ProDom; PD000650; Glyco\_hydro\_1; 1.  
 DR Signal.  
 FT SIGNAL  
 KW  
 FT CHAIN  
 SQ SEQUENCE 434 AA; 48363 MW; 121ABE2E3AF2B7C CRC64;

Query Match 39.4%; Score 1042; DB 3; Length 434;  
 Best Local Similarity 47.6%; Pred. No. 2,4e-75;  
 Matches 224; Conservative 59; Mismatches 140; Indels 48; Gaps 13;

QY 7 LPNDFEMGFATAAQTGEGGGRPSIMDTYCHLEPSRTNGANGVACDHRYDEDF 66  
 DB 2 LPNDLQGFKAQAQTGEGGGRPSIMDTFCAGCKTADSSGVYACDSTNRTAEDI 61  
 QY 67 DLTLYGAKAYRFSLSWSRIIPLCGRDLPVNEEGIEFYSKILDALRRGITPWTLYHMD 126  
 DB 62 ALKSLGAKSYRFSIS-SR-IPESGRDANQAGIDHYVAFVDLLAGITPFTTLPHMD 119  
 QY 127 LPQALHRYGGMVNEEVQDPERYARLCERFGDRVQNNITNXPMIOAIYGATGSNA 186  
 DB 120 L--LHQRYGGLNRTFPLDFENYARVMF-RALPKVRNM---NEPLCSAIPGVSGSFA 172  
 QY 187 PGSSINKESTEGTATEPMLAGKAQIMSHARAAYVSRDPR-SQCGQIGISLNGEY 245  
 DB 173 PGROS-----TSEPWTGNNILVAHGRAVAKAYRDFFKPSGGQIGIVLNGDFTY 222  
 QY 246 PMDSNEPRDKEAERMEFHIGFANPIFLKQDYPESMKKQGERLPALTPADPAIINAG 305  
 DB 223 PMDAADPADKE-----RLFEFTAMPADPIYL-GDYPASMRKQDRLFTTPEERALVH-G 276  
 QY 306 ETDPFYGMNVTTSQFARHLDPVPETDYLAGA-IHEHOKKDGSPVGEESGLAMLRSCPMPFR 365  
 DB 277 SNDFYGMNHTSYNIRHRSBPASADDTVGNVDVLFTHKQNCIGIPETQSPMLRCAGAFR 336  
 QY 366 KHLARVGLYGK-PIYITENGCCPGSENNTCBAVNDPRIRYFSDHLSISAKAITQDG 424  
 DB 337 DFLVMTSKRYGSPPIYITENGTSIKGSSDLPEKXILDDPRVKKYNYEIAMVAVAVELDG 396  
 QY 425 VVVGKGFAMALLDNLEWSDGVRFGVTFDTYTT-LKRTPKKSLVLCQMF 474  
 DB 397 VNV-----RFGVTVVDYENGQRFRPKSKSLKPLF 427

RESULT 9  
 Q9FIM4 PRELIMINARY; PRT; 490 AA.  
 AC Q9FIM4;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)  
 DE Beta-glucosidase.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=9915623; PubMed=10048488;  
 RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.  
 RT Sequence features of the regions of 1,081,958 bp covered by seventeen  
 RT physically assigned P1 and TAC clones."  
 RL DNA Res. 5:379-391(1998).  
 DR EMBL; AB016877; BAB11630.1; -.  
 DR HSP; P26205; ICBG.  
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001360; Glyco\_hydro\_1.  
 DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS; PR00131; GHYDRLASE1.  
 DR ProDom; PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 SQ SEQUENCE 490 AA; 56077 MW; 50DEB68FBF06D538 CRC64;

Query Match 39.2%; Score 1037.5; DB 10; Length 490;  
 Best Local Similarity 45.0%; Pred. No. 6,6e-75;  
 Matches 209; Conservative 83; Mismatches 155; Indels 17; Gaps 9;

QY 8 PNDPEWGFATAAQTGEGGGRPSIMDTYCHLEPSRTNGANGVACDHRYDEDF 67  
 DB 22 PSFTFTGVATSAQTGEGGGRPSIMDKFTHIEGKILDGSDNGVADHYHRYKEDVD 81  
 QY 68 LLTRYGAKAYRFSLSWSRIIPLCGRDLPVNEEGIEFYSKILDALRRGITPWTLYHMD 127  
 DB 82 LIQQLGFGAYRFSISWSRIIP-DGLGTEVNEEGIAFYNDLINTLLEKGIOPYVTLYHMDL 140  
 QY 128 PQALHRYGGMVNEEVQDPERYARLCERFGDRVQNNITNXPMIOAIYGATGSNA 187  
 DB 141 PSHLQEAIGGWTNRKIVDY-FGLYADACFANFDRVHMTTLNEPIQTSVNGCIGIFAP 199  
 QY 188 GRSSINKESTEGTATEPMLAGKAQIMSHARAAYVSRDPRSPQCGQIGISLNGEYEP 247  
 DB 200 GRNE-----KPLIEPIVSHHQLAHATANSIRSKYKESQCGQIGISVCEWMEP- 250  
 QY 248 DSNEPRDKEAERMEFHIGFANPIFLKQDYPESMKKQGERLPALTPADPAIINAGET 307  
 DB 251 NSKPEEDKVAADRIRIDFGQMFDPPLFF-GDYPASMRKQDRLFTTPEERALVH-G 309  
 QY 308 DPFYGMNVTTSQFARHLDPVPETDYLAGA-IHEHOKKDGSPVGEESGLAMLRSCPMPFR 365  
 DB 310 DFLGLNHTYSRLISHVSNKEAESNFYQAOELERIVLEENDLIGERASDMLAVPWGIR 369  
 QY 366 KHLARVGLYGK-KPIYITENGCCPGSENNTCBAVNDPRIRYFSDHLSISAKAITQDG 424  
 DB 370 KTLNYSKRYNHPPIITENGMDDEDDGASIHMDLDDKRRVYFYFSYLANVQAI-EDG 428  
 QY 425 VVVGKGFAMALLDNLEWSDGVRFGVTFDTYTT-LKRTPKKSA 467  
 DB 429 VDIKGFAMSLDNLEWAGYTRFGVYVVDYKNGLTRHPSKA 472

RESULT 10  
 Q8LF56 PRELIMINARY; PRT; 498 AA.  
 AC Q8LF56;  
 DT 01-OCT-2002 (TReMBLrel. 22, Created)  
 DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)  
 DT 01-JUN-2003 (TReMBLrel. 24, last annotation update)  
 DE Beta-glucosidase, putative.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;



Db 273 FEB-ESKKTEDIEAQAQADPQGLWFLDPLMF-GDYSSSMKSRVGRSLPVFTGSQSSLVK 330  
QY 304 AGETDFGNNYTSQFARHLDPVPEFDYLGAIHEHOENKDG-----SPGEESSGL 354  
Db 331 -GSLDFVGINHYYTYVARN-----NATNLGTLHDAVSSGTYTLPEFKLSTIGDAASS 384  
QY 355 AMLRSCPMFRKHLARVGLYK-PYITENGCPGCEENMTCEAVNDPFRIRYFDSHL 413  
Db 385 IMYIYVGRKSLMNNYKHXRGPNPVITENGMDPNSILSRDALKDAKRIKHYHDL 444  
QY 414 DJSKATODGVVVGKFAWALLDNLMSGCGRGVFTDY-TLKRTPKKS 466  
Db 445 SSLQASIKEDGCVNGKGFVMSLLDNEMWAGYSSRFLYFVDYRDNKRYPKDS 498

## RESULT 12

QY073 PRELIMINARY; PRT; 489 AA.  
AC QY073; (TEMBLrel. 25, Created)  
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
DE Latex cyanogenic beta glucosidase.  
OS Hevea brasiliensis (Para rubber tree).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid 1; Malpighiales; Euphorbiaceae; Crotonoideae; Mitrandaeae;  
OC NCBI\_TaxID=3981;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sunderaasan S., Yeang H.Y.;  
RT "Hevea brasiliensis latex cyanogenic glucosidase (Hev b 4 a major latex allergen)".  
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY297039; AAPS1059.1; -  
SQ SEQUENCE 489 AA; 56132 MW; 1D79CE651976FB94 CRC64;

Query Match 38.9%; Score 1028.5; DB 10; Length 489;  
Best Local Similarity 45.9%; Pred. No. 3.5e-74;

Matches 214; Conservative 83; Mismatches 146; Indels 23; Gaps 13;

QY 8 PNDPFWGFATAAYOIEGAVKEGGRGSPIMDYCHLEPSRTNGANGDVA CDHYHRDEDFD 67  
Db 24 PNFPLRGVAISAQIEGCGREGGRGSIWAFSHTKTILDGSGDVAVHYHRYKEDIE 83  
QY 68 LTTYGAKAYRFLSWSRIIP-LGGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHW 125  
Db 84 LIAKLFDAVRFSLSMSRIIPDGLGTR--VNEGIAFYNNIINALLEKGIPEYITLYHW 140  
QY 126 DLQALHDRVGGMLNBEVQDLDFERYALCFERFGDGVQWMTINXPMIOAIYGATGSN 185  
Db 141 DLPLHLQESNGWLNKEIVRY-PAIYADTCFASFGGRVKKMITLNEPLQTAVANGPFDGIL 199  
QY 186 APGRSSINXKSTGNTATEPMLAGKQIMSHARAVAYSRDPFSOKGOIGISLNGDYE 245  
Db 200 APGR---HEIS-----YTFPLASHHQIILAHATAVSIYRSMTYDNQGGVGLVDDCEWAB 251  
QY 246 PWDNSNEPRDEKAERMEFHIWGPANDIFLKQY PESHKQOLGERLPALTPADPAIINAG 305  
Db 252 S-NSDKIEDKAAAKRLEFQGLWYHPLYY-GDYPERMKRIKGGGLPKPSEEDKELLR-N 308  
QY 306 EYDFGMYNTYTSQFAHLDGPVETDYLGA--IHEHOENKDSGPGVEESGLALRSCPD 363  
Db 309 SLDFIGLNHYSRFRKGVTDSPACYYKQAEIERLAKWDEGPIGERAASEMLYYRPGW 368  
QY 364 FRGHLARVGLYKGP-IYITENGCPGCEENMTCEAVNDPFRIRYFDSHLSKAITO 422  
Db 369 LRKLVNIYVGRVNNPIIYITENGMD-DESSAPLHEMLDKLAVRIFKGLAVALVAQAI-K 426  
QY 423 DGAVVKGYPAMALLDNLMSDGYGPRGVTFTDYTT-LKRTPKKSA 467

Db 427 DGADVGRYFAMSLLDNEMWAGYTKRFGILYVDYKNGLARHPKSSA 472

## RESULT 13

QY0406 PRELIMINARY; PRT; 515 AA.  
AC QY0406;  
DT 01-DEC-2001 (TEMBLrel. 19, Created)  
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
DE Putative beta-glucosidase.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Eriarthoideae; Oryzaceae; Oryza.  
OC NCBI\_TaxID=4530;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nipponbare;  
RA Nascimento L., Spiegel L., de la Bastide M., Kirchhoff K., King L.,  
RA Preston R., Vil M.D., Baker J., Zucavern T., Santos L., Bell M.,  
RA Miller B., Kuit K., Rodriguez S., Cunniff D.M., Balija V., Shah R.,  
RA Bahre A., Palmer L., Yang C., O'Shaughnessy A., Dedhia N.,  
RA McCombie W.R.;  
RT "Genomic Sequence for Oryza sativa, Nipponbare strain, clone  
RT OJNBA0065C16, from chromosome 10, complete sequence."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC074354; MAK92581.1; -  
DR Gramine; Q94H06; -  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001360; Glyco\_hydro\_1.  
DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
DR PRINTS; PR00131; GLYHTRLSL.  
DR ProDom; PD000650; Glyco\_hydro\_1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
SQ SEQUENCE 515 AA; 58683 MW; B842BD29627AE8CC CRC64;

Query Match 38.7%; Score 1023; DB 10; Length 515;  
Best Local Similarity 42.3%; Pred. No. 1e-73;

Matches 208; Conservative 88; Mismatches 160; Indels 36; Gaps 12;

QY 6 ALPNDPFWGFATAAYOIEGAVKEGGRGSPIMDYCHLEPSRTNGANGDVA CDHYHRDEDFD 65  
Db 37 SPFNGVFPGRASSAYIEGAVKEGGRGPTIWDKFAITPKKIIDFSNADVAVQYHFEED 96  
QY 66 FDLITYGAKAYRFLSWSRIIP-LGGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHW 125  
Db 97 IQLMADMGMDAVRFSISMSRIIPNG--TGEVQAQIDHYNKLINALLAGIPEYITLYHW 154  
QY 126 DLQALHDRVGGMLNBEVQDLDFERYALCFERFGDGVQWMTINXPMIOAIYGATGSN 185  
Db 155 DLQALHDRVGGMLNBEVQDLDFERYALCFERFGDGVQWMTINXPMIOAIYGATGSN 213  
QY 186 APGRSSINXK-STEGNTATEPMLAGKQIMSHARAVAYSRDPFSOKGOIGISLNGDYE 243  
Db 214 APGRSGVHLHLYCKKNSGTEPIYVAHMTLSHATYSOIYRKKYKASONGEIGISDVW 273  
QY 244 YEPWDSNEPRDEKAERMEFHIWGPANDIFLKQY PESHKQOLGERLPALTPADPAIINAG 303  
Db 274 YEPW---SNSYADIEAARQAQDFQGFADPFF-GDYPATMRSRVGRSLPKPTEKEALVN 331  
QY 304 AGETDFGMYNTYTSQFAHLDGPVETDYLGA--IHEHOENKDSGPGVEESGLALRSCPD 363  
Db 332 -GSLDFWGINHYTFTTKDQSTVIEKLLNLTADPATISV-----PFRNGOPIGIRANS 385  
QY 355 AMLRSCPMFRKHLARVGLYKGP-IYITENGCPGCEENMTCE-----EAVNDPF 404  
Db 386 IMYIYVGRKSLMNNYKHXRGPNPVITENG-KC---TYVCDLFLPISIKMLAKDK 441  
QY 405 RIRYFDSHLSISKAITODGVVVGKYPAMALLDNLMSDGYGPRGVTFTDYTT-LKRTPK 464  
Db 442 RTKYHNDYLTNLADSIREDGCVRGYFAMSLLDNEMWAGYTSRGLYVDYKNGRKRPK 501



Db           491   :::|  
              KKEIRV 496

Search completed: March 25, 2004, 15:56:36  
Job time : 48 secs

---

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2004, 15:45:56 ; Search time 17 Seconds

(without alignments)  
1482.468 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644

Sequence: 1 MPESLALPNDENGFATAY.....KSALVTKDMPARQVKA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	880.5	33.3	493	1	BGLS_TRIRP
2	879	33.2	566	1	BGLC_MAIZE
3	858.5	32.5	548	1	MYRO_BRANA
4	857.5	32.4	1927	1	LPH_HUMAN
5	852.5	32.2	1936	1	LPH_RABIT
6	851.5	32.2	1928	1	LPH_RAT
7	849.5	32.1	541	1	MYRO_ARATH
8	845	32.0	544	1	MYR3_SINLA
9	839	31.7	446	1	BGLA_THEMA
10	831	31.4	528	1	BGL_ARATH
11	809	30.6	444	1	BGLA_THENE
12	802	30.3	449	1	BGLA_BACCI
13	791	29.9	448	1	BGLB_PABPO
14	790	29.9	425	1	BGLT_TRIRP
15	789	29.8	448	1	BGLA_PABPO
16	767	29.0	459	1	BGLS_AGRSP
17	749	28.3	448	1	BGLA_CLOTH
18	729.5	27.6	455	1	BGLS_CALSA
19	702	26.6	477	1	BGL2_BACSU
20	689.5	26.1	473	1	BGLB_MICBI
21	614	23.2	468	1	BGLG_LACIA
22	607.5	23.0	474	1	LACG_LACCA
23	599	22.7	470	1	LACG_STANU
24	597	22.6	470	1	LACG_STAM
25	590.5	22.3	473	1	LACG_LACAC
26	587	22.2	468	1	LACG_STAMU
27	583	22.0	470	1	LACG_STAP
28	558.5	21.1	480	1	BGLA_ERWHE
29	530.5	20.1	479	1	BGLA_ERWHE
30	522.5	19.8	479	1	BGLA_ERWHE
31	497	18.6	473	1	ABGA_CLOLO
32	491	18.6	464	1	CASB_XLBEX
33	485	18.3	470	1	BGLB_ECOLI

34	475.5	18.0	474	1	ASCB_ECOLI
35	474.5	17.9	469	1	BGLI_BACSU
36	470	17.8	465	1	ARAB_ERWCH
37	358.5	13.6	489	1	BGLA_SULSO
38	344.5	13.0	489	1	BGLA_SULSO
39	343.5	13.0	491	1	BGLA_SULAC
40	304.5	11.5	243	1	MYR2_SINLA
41	302	11.4	248	1	MYR2_SINLA
42	301.5	11.4	244	1	MYR1_SINLA
43	97.5	3.7	1088	1	YMS_YEAST
44	96.5	3.6	766	1	YMS_YEAST
45	96	3.6	876	1	SYL_RHIME

## ALIGNMENTS

RESULT 1	ID	BGLS_TRIRP	STANDARD	PRT	493 AA.
AC	P26204				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	01-FEB-1996 (Rel. 33, Last annotation update)				
DE	Non-cyanogenic beta-glucosidase precursor (EC 3.2.1.21)				
OS	Trifolium repens (Creeping white clover)				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; rosids;				
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.				
NCBI	_TaxID=3899;				
RN	[1]				
RP	SEQUENCE FROM N.A. AND SEQUENCE OF 135-157.				
RC	SPRAIN=5100 (EC); TISSUE=leaf;				
RX	MEDLINE=91322517; PubMed=1907511;				
RA	Oxley E., Dunn M.A., Pancoro A., Hughes M.A.;				
RT	"Nucleotide and derived amino acid sequence of the cyanogenic beta-				
RL	glucosidase (linamarase) from white clover (Trifolium repens L.).";				
PL	Plant Mol. Biol. 17:209-219(1991).				
CC	- CARBOLYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-				
CC	glucose residues with release of beta-D-glucose.				
CC	- TISSUE SPECIFICITY: Leaves.				
CC	- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL; X56734; CAA0058.1; ..				
DR	PIR; S16581; GLY31.				
DR	HSSP; P26205; 1CBG.				
DR	InterPro; IPR001360; Glyco_hydro_1.				
DR	Pfam; PF00232; Glyco_hydro_1; 1.				
DR	PRINTS; PR00131; GLYTRASE1.				
DR	PRODOM; PD000650; Glyco_hydro_1; 1.				
DR	PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.				
DR	PROSITE; PS00573; GLYCOSYL_HYDROL_F1_2; 1.				
KW	Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	STGNLU	1	18		POTENTIAL.
FT	CHAIN	19	493		NON-CYANOGENIC BETA-GLUCOSIDASE.
FT	ACT_SITE	204	204		PROTON DONOR (POTENTIAL).
FT	ACT_SITE	422	422		NUCLEOPHILE (BY SIMILARITY).
FT	CARBOHYD	34	34		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	335	335		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	371	371		N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	412	412		N-LINKED (GLCNAC...) (POTENTIAL).
SQ	SEQUENCE	493 AA;	55960 MW;	B6B3BA5B8F8C8F	CRG4;

Query Match 33.3%; Score 880.5; DB 1; Length 493;

Best local Similarity 40.3%; Pred. No. 8,2e-61;

```
Matches 186; Conservative 80; Mismatches 177; Indels 19; Gaps 8;
QY 6 ALPNDEMGFATAAYOIEGAVKGGKPSIMPTYCHLEPSR-TNGANGDVACHHHRDE 64
D 39 SFFRGRIFAGSSAVYFEGAVNEGGRGFSIMTFTFKKPEKTRDGSNADITVDQYHRKE 98
QY 65 DFDLTKYAKAYRFSLSMSRIIPUGRLDPVNEEGIEFYSKLDLALRGITPTWLYH 124
D 99 DVGIMDKQMDSYRFSISWPRILPKGKLSGINHEGIKYNNLWELLANGIQPVLTFH 158
QY 125 WDLPOLAHRYGGMVNEVQDLDFERYARLCEFRGDRQWMTTNNXPQAIYATGS 184
D 159 WDLPOLLEDEYGGFLNSGVIN-DFRDYDULCEKEGDRVRVYSTNEPWFNSGALGT 217
QY 185 NAGRSSINKHSTEGNTATEPWLAKAQMISHARAVAVSRDPSRQKQIGISLNGDY 244
D 218 NAGRCASANVAKPGDSSGTGYIVHNOILAHAVHYKTYQAYQKGIITLVSNWL 277
QY 245 EFWDSNEPDKAERREFFHIGWPNPIFLKKDYPESSKQGLRPLATPADPAIINA 304
D 278 MFLDONSIPDIKAERSLDFQGLFMEQL-TTGDYSKSMRIRIVKRLPKFSFESSLVN- 335
QY 305 GETDYGMYNYSOAFBHLDG-----PVPEYDYLGAHHEQENKQGSVYGEESGLAMLR 359
D 336 GSPFDIGINYSSSTISNAPSHGNAKPSYSTNPMNTI--SFEKRGIFLGPRAASIMTYV 392
QY 360 CPDMFRKHLARYVLYGK-----PIYITENGCPGCEENMTCEAVNDPRIRYFDSHL 413
D 393 YPMYFIOGDEFIFCYILKINLTITLOFSITENGMENFNATLPVEBALNTYRIDYVYHL 452
QY 414 DSISAIQTODGVVVGCFAMALLDLNLEMSDGYGPRFGVTFTD 455
D 453 YYIRSAI-PAGSNVKGAFYAMSLDCNEMFAGTFVRGLNFVD 493

RESULT 2
BGLC_MAIZE
ID_BGLC_MAIZE STANDARD; PRT; 566 AA.
AC P49235;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Beta-glucosidase, chloroplast precursor (EC 3.2.1.21) (Gentiobiase)
DE (Cellubiase) (Beta-D-glucoside gluconhydrolase).
GN GLUT1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=CV. INHERD LINE K55; TISSUE=Shoot;
RA Eesen A., Shahid M.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RC STRAIN=CV. MUTIN; TISSUE=coleoptile;
RX MEDLINE=94053747; PubMed=8235622;
RA Biziohady B., Moore I., Kristoffersen P., Bako L., Campos N.,
RA Schell J., Palme K.;
RT "Release of active cytokinin by a beta-glucosidase localized to the
RT maize root meristem";
RL Science 262:1051-1054(1993).
RN [3]
RC SEQUENCE OF 55-74, AND CHARACTERIZATION.
RC STRAIN=CV. INHERD LINE K55;
RA Eesen A.;
RT "Purification and partial characterization of Maize (Zea Mays L.)
RT beta-glucosidase";
RL Plant Physiol. 98:174-182(1992).
RN [4]
RC SEQUENCE OF 55-69; 165-174; 207-213 AND 217-235.
```

```
RC TISSUE=Coleoptile;
RA Touzet P., Ricciardi F., Morin C., Demerval C., Huet J.-C.,
RA Fernollet J.-C., Zavy M., de Vienne D.;
RT "The maize two dimensional gel protein database: towards an integrated
RT genome analysis program";
RL Theor. Appl. Genet. 93:997-1005(1996).
CC -1- FUNCTION: Is implicated in many functions such as ABA metabolism,
CC hydrolysis of conjugated gibberellins, conversion of storage forms
CC of cytokinins to active forms. Also acts in defense of young plant
CC parts against pests via the production of hydroxamic acids from
CC hydroxamic acid glucosides. Enzymatic activity is highly
CC correlated with plant growth. Its optimal pH is 5.8, and optimal
CC temperature 50 degrees Celsius. Activity is totally lost when the
CC enzyme is heated at 55 degrees Celsius.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC glucose residues with release of beta-D-glucose.
CC -1- ENZYME REGULATION: Reversibly inhibited by micromolar
CC concentrations of Hg(2+) or Ag(+), but irreversibly inhibited by
CC alkylation in presence of urea.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- TISSUE SPECIFICITY: Most abundant in the coleoptile.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U25157; AAA65946.1; -.
DR EMBL; X74217; CAAS2293.1; -.
DR PIR; A48860; A48860.
DR PDB; 1E1E; 19-FEB-01.
DR PDB; 1E1F; 19-FEB-01.
DR PDB; 1E4L; 11-DEC-00.
DR PDB; 1E4N; 11-DEC-00.
DR PDB; 1E55; 11-DEC-00.
DR PDB; 1E56; 11-DEC-00.
DR Maize-2DPAGE; P49235; COLEOPTILE.
DR MaizeDB; 13870; -.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GHYDRASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolyase; Glycosidase; Chloroplast; Transit peptide; 3D-structure.
FT TRANSIT 1
FT CHAIN 55
FT ACT SITE 245
FT ACT SITE 460
FT CONFICT 477
FT CONFICT 551
FT CONFICT 554
SQ SEQUENCE 566 AA; 64237 MW; 4EA241258AE3641B CRC64;

Query Match 33.2%; Score 879; DB 1; Length 566;
Best Local Similarity 38.3%; Pred. No. 1.3e-60;
Matches 192; Conservative 94; Mismatches 183; Indels 32; Gaps 14;

QY 1 MPESIALPNDFMGFATAAYOIEGAVKGGKPSIMPTYCHLEPSR-TNGANGDVACHY 59
D 72 IPQRWPFSPDFTFGAATSAVOIEGAMNDGKESWMDHFCNHPRIIRIDSGNSY 131
QY 60 HRYDEDFDLTKYAKAYRFSLSMSRIIPUGRLDPVNEEGIEFYSKLDLALRGITPW 119
D 132 HMYKTDVRLKEMGMDAVRFSISWPRILPKGKEGGINPDGIKYRNLINLLENGIRPY 191
QY 120 VTLVHMDLPOLAHRYGGMVNEVQDLDFERYARLCEFRGDRQWMTTNNXPQAI 177
D 192 VTLFHWMDLPOLAEKRYGGLDKSHKSIIVEDYTYFAKCFDNFGDKVKNMLTENEPQFTS 251
```



```

QY 178 YGVATGSAAPRRS-- -INXSTEGENTTEPEMLACKAQMASHARAVAVYSDPFRSQGOI 235
Db 252 FSYIGVGFAPRCSPGDLCAVPTGNSLVEPTTAGHNILAHAEVNDLYNKAYKDDT-RI 310
QY 236 GISLNGDYEEWDSNEPRDKEAERERMEFHIGWEFANPIFLKKDYPESNKKOLGERPALT 295
Db 311 GLARDVWGAVPYGRSF- LDKQAERMSMDINLGWLEPV- VARGDVPFSMRSLARERLLPFFK 368
QY 296 PADPAILINAGETDFYGMANNYSQPARHLDD----GVPETDYLGAIEHGOENKGSPPGE 350
Db 369 DEQEKEL-AGSYNNMLGLNYYTSRFSKNIDIPNYSPLVNTDADAYASOE- VNGPDGKPIGP 426
QY 351 ESGLAWLRSQCDMPKRLHARYGLYGYK-PIYITENG-----CPCGGEENMTCEEAVN 401
Db 427 PMGPMPIYMMPEGLDGLMLTKMKRYGNPFIITENGIDVDUTKETPLD-----MEALIN 480
QY 402 DPFPIRYFDSHLSDISKAITODGVVKKGYFAMALLDLNLEWSDGYGPREGVFTFD- YTTLK 460
Db 481 DYKRLDVIQRIIATLKESIDL- GSNVQGYFAMSLDNEFWAGFTERYGIYVDRNNCT 539
QY 461 RTPPKSALVTKOMFARORVK 481
Db 540 RYMKESAKWLKEFNTAKKPSK 560

RESULT 3
MYRO_BRANA
AC 000326;
ID MYRO_BRANA STANDARD; PRT; 548 AA.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myrosinase precursor (EC 3.2.1.147) (Sintgrinase) (Thioglucosidase).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OC NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Svalofe Karat;
RA Falk A., Xue J., Jenman M., Raak L.;
RT "Sequence of a cDNA clone encoding the enzyme myrosinase, and
expression of myrosinase in different tissues of Brassica napus.";
RL Plant Sci. 83:181-186(1992).
CC -1- FUNCTION: Degradation of glucosinolates (glucose residue linked by
a thioglucoside bound to an amino acid derivative) to glucose,
sulfate and any of the products: thiocyanates, isothiocyanates,
nitriles, epithionitriles or oxazolidine-2-thiones.
CC -1- CATALYTIC ACTIVITY: H2O + a thioglucoside = a sugar + a thiol.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Vacuolar.
CC -1- TISSUE SPECIFICITY: In vacuoles called myrosin grains of a certain
class of cells, myrosin cells, distributed in the cotyledons and
the axis of the embryo as well as in different organs of the
growing plant.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/or\_send\_an\_email\_to\_license@isb-sib.ch).
CC -----
DR EMBL; X60214; CAA42775.1; -
DR PIR; S26149; S26149.
DR HSSP; P26205; 1CRG.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR

```

Query Match	32.5%	Score 858.5	DB 1	Length 548
Best Local Similarity	40.1%	Pred. No. 4.8e-59		
Matches 194	Conservative 87	Mismatches 156	Indels 47	Gaps 16
Dr	Prodom: PD00650; Glyco hydro 1; 1.			
DR	PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.			
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.			
KW	Hydrolase; Glycosidase; Multigene family; Glycoprotein; Signal.			
FT	SIGNAL	1	20	BY SIMILARITY.
FT	CHAIN	21	548	MYOSINASE.
FT	ACT SITE	429	429	NUCLEOSIDE (BY SIMILARITY).
FT	CARBOHYD	110	110	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	240	240	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	548 AA	62735 MW	F30346A01CD732A CRC64;
Query Match	32.5%	Score 858.5	DB 1	Length 548
Best Local Similarity	40.1%	Pred. No. 4.8e-59		
Matches 194	Conservative 87	Mismatches 156	Indels 47	Gaps 16
Dr	Prodom: PD00650; Glyco hydro 1; 1.			
DR	PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.			
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.			
KW	Hydrolase; Glycosidase; Multigene family; Glycoprotein; Signal.			
FT	SIGNAL	1	20	BY SIMILARITY.
FT	CHAIN	21	548	MYOSINASE.
FT	ACT SITE	429	429	NUCLEOSIDE (BY SIMILARITY).
FT	CARBOHYD	110	110	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	240	240	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	520	520	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	548 AA	62735 MW	F30346A01CD732A CRC64;

RT		"Complete primary structure of human and rabbit lactase-phenylalanine
RT		hydrolase; implications for biosynthesis, membrane anchoring and
RT		evolution of the enzyme.";
RL		EMBO J. 7:12705-2713(1988).
RN	(2)	
RP		SEQUENCE FROM N.A.
RX		MEDLINE=91206402; PubMed=1902057;
RA		Boll W., Wagner P., Mantel N.;
RA		"Structure of the chromosomal gene and cDNAs coding for lactase-
RT		phenylalanine hydrolase in humans with adult-type hypolactasia or
RT		persistence of lactase.";
RL		Am. J. Hum. Genet. 48:889-902(1991).
CC	-1-	FUNCTIOIN: LPH splits lactose in the small intestine.
CC	-1-	CATALYTIC ACTIVITY: Lactose + H(2)O = D-glucose + D-galactose.
CC	-1-	CATALYTIC ACTIVITY: Glycosyl-N-acetylsphingosine + H(2)O = a sugar +
CC		N-acetylsphingosine.
CC	-1-	SUBCELLULAR LOCATION: Type I membrane protein. Brush border.
CC	-1-	TISSUE SPECIFICITY: Intestine.
CC	-1-	DOMAIN: THE SEQUENCE EXHIBITS 4 REGIONS (I-IV) OF INTERNAL
CC		HOMOLOGY; THEREFORE LPH MIGHT HAVE EVOLVED BY TWO CYCLES OF
CC		PARTIAL GENE DUPLICATION.
CC	-1-	DISEASE: Defects in the upstream vicinity of LCT are the cause of
CC		disaccharide intolerance III [MIM:223100], also known as adult
CC		lactase deficiency or adult-type hypolactasia. In many human
CC		populations the activity of LCT declines in adults, leading to
CC		adult-type hypolactasia, whereas in other populations the high
CC		activity persists.
CC	-1-	SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC		or send an email to licenses@isb-sib.ch).
CC		-----
DR	EMBL,	X07994; CAA30801.1; .
DR	EMBL,	M61850; AAA59504.1; JOINED.
DR	EMBL,	M61834; AAA59504.1; JOINED.
DR	EMBL,	M61835; AAA59504.1; JOINED.
DR	EMBL,	M61836; AAA59504.1; JOINED.
DR	EMBL,	M61837; AAA59504.1; JOINED.
DR	EMBL,	M61838; AAA59504.1; JOINED.
DR	EMBL,	M61839; AAA59504.1; JOINED.
DR	EMBL,	M61840; AAA59504.1; JOINED.
DR	EMBL,	M61841; AAA59504.1; JOINED.
DR	EMBL,	M61842; AAA59504.1; JOINED.
DR	EMBL,	M61843; AAA59504.1; JOINED.
DR	EMBL,	M61844; AAA59504.1; JOINED.
DR	EMBL,	M61845; AAA59504.1; JOINED.
DR	EMBL,	M61846; AAA59504.1; JOINED.
DR	EMBL,	M61847; AAA59504.1; JOINED.
DR	EMBL,	M61848; AAA59504.1; JOINED.
DR	EMBL,	M61849; AAA59504.1; JOINED.
DR	PIR,	S01168; S01168.
DR	HSSP,	P26205; ICBG.
DR	Gneww,	KGNC:6530; LCT.
DR	MIW,	603202; .
DR	MIW,	223100; .
DR	GO:	GO:0005887; C:integral to plasma membrane; TAS.
DR	GO:	GO:0005624; C:membrane fraction; TAS.
DR	GO:	GO:0000016; F:lactase activity; TAS.
DR	InterPro:	IPRO01360; Glyco_hydro_1.
DR	Pfam:	PF00232; Glyco_hydro_1; 4.
DR	PRINTS:	PR00131; GLHYDRLASE1.
DR	ProDom:	PD000650; Glyco_hydro_1; 4.
DR	PROSITE:	PS00572; GLYCOSYL_HYDROL_F1_1; 2.
DR	PROSITE:	PS00653; GLYCOSYL_HYDROL_F1_2; 3.
KW	Hydrolase;	Glycosidase; zymogen; Signal; Transmembrane; Repeat.
FT	SIGNAL	1 19
FT	PROPEP	20 866 BETA-GLUCOSIDASE.
FT	CHAIN	867 1927 LACTASE-PHENYLALANINE HYDROLASE.

FT	DOMAIN	20	1982	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1883	1901	POTENTIAL.
FT	DOMAIN	1902	1927	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	87	1841	4 X APPROXIMATE REPEATS.
FT	REPEAT	87	172	1.
FT	REPEAT	363	848	2.
FT	REPEAT	884	1365	3.
FT	REPEAT	1370	1841	4.
FT	ACT SITE	1065	1065	PROTON DONOR (POTENTIAL).
FT	ACT SITE	1273	1273	NUCLEOPHILE (BY SIMILARITY).
FT	ACT SITE	1538	1538	PROTON DONOR (POTENTIAL).
FT	ACT SITE	1749	1749	NUCLEOPHILE (BY SIMILARITY).
FT	CONFLICT	219	219	V -> I (IN REF. 2).
FT	CONFLICT	1639	1639	N -> S (IN REF. 2).
SO	SEQUENCE	1927 AA;	218601 MW;	FA48640B9A35A6F CRC64;
Query Match 32.4%; Score 857.5; DB 1; Length 1927;				
Best Local Similarity 40.3%; Pred. No. 3e-58;				
Matches 192; Conservative 75; Mismatches 167; Indels 43; Gaps 12;				
Qy	8	PNDPEMGAPATAVOIEGAVNEGGRSPIMPTJCHLEFSRTNGANGDVACDHRYHDDDFD	67	
Db	1378	PEGFTWSPASAAVAOIEGAMRADGKSLTWPFTSTPLRVENDAIQDVACSHYAIADLV	1437	
Qy	68	LITRYGAKAVRFSLSWSRIIPLGRLDPVNEEGIEFYSKLIDALLRGTTPWTLYHMDL	127	
Db	1438	TLQMLGVSHRFSSISWGRILP-DGTRYINIDAGLNTYVRLIDTLAASIQPQVLIYHMDL	1496	
Qy	128	PQALHRYGGMWLNVEQLDPERFARLCEFRFGDROVQNTITXPMIQAIIYGAATGSNAP	187	
Db	1497	PQTIQD-VGGMEN-ETTVORPKETADVLQRLGDKVFWITLKEPFVIAVQGYGYTAAP	1554	
Qy	188	GRSSINGHSTEGNATEPMLAGKAQIMSHARAAVAVSRDRPSQGOIGISLNDYYEPW	247	
Db	1555	GVSN-----RGTAPYIVGHNLIKAHAEMLVNDVYRAQSGVISTISSDMAEPR	1606	
Qy	248	DSNEPRKEAERMEFHIGWNPILKKDYPESMKQIGF-----RIPALTPA	297	
Db	1607	DPSQOEVEAEARRYVQPMGWFAPHPIKNDYNEVMKTRLRDRSLAGLNKSRLPEFTES	1666	
Qy	298	DFALINAGEIDFQVMNYYISOPARHLIDGVPENDYLGAIHEHQNKQSPVGE-----ESG	353	
Db	1667	EKRIN-GTYDFRFRNHYTTLVLA--NATALLSPADRGVAISADBSWPDSC	1718	
Qy	354	LAWLRSCDPMFRKHLARVGLYG-KPIYITENGCPCEBENMTCEAVNDPFRIRYDISH	412	
Db	1719	SFWIKMTPFGFRRLRLMLKEEYNDPPIYVENG--SQBETDLNLTARIYULRY	1772	
Qy	413	LDSISKATIDDGVVVKGIFYAMALLDNLSEWSDGGPRGVTFTDYT--TLKRTPKSA	467	
Db	1773	INEALKAV-ODKVDLRGYTWSANDNEFMAIGSERFGELHFVNISDPSLPRIIPASA	1828	
RESULT 5				
ID_LPH_RABIT STANDARD; PRT; 1926 AA.				
AC	P08649;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	Lactase-phenolizin hydrolase precursor (lactase-glycosylceramidase)			
DE	[Includes: Lactase (EC 3.2.1.108); Phenolizin hydrolase (EC 3.2.1.62)].			
GN	ICT OR LPH.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID=9986;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.			
RC	STRAIN=New Zealand white.			
FX	MEDLINE=8903064; Pubmed=2460343;			
ZA	Mantel N., Villa M., Nuzler T., Wacker H., Boll W., James P., Hunziker W., Semenza G.,			

RT "Complete primary structure of human and rabbit lactase-phlorizin  
 RT hydrolase: implications for bioynthesis, membrane anchoring and  
 RT evolution of the enzyme."  
 RL EMBO J. 7:2705-2713(1988).  
 RN  
 RP ACTIVE SITES.  
 RX MEDLINE=92406791; Pubmed=1388157;  
 RA Wecker H., Keller P., Falchetto R., Legler G., Semenza G.;  
 RT "Location of the two catalytic sites in intestinal lactase-phlorizin  
 RT hydrolase. Comparison with sucrose-isomaltase and with other  
 RT glycosidases, the membrane anchor of lactase-phlorizin hydrolase."  
 RL J. Biol. Chem. 267:18744-18752(1992).  
 CC -1- FUNCTION: LPH splits lactose in the small intestine.  
 CC -1- CATALYTIC ACTIVITY: Lactose + H(2)O = D-glucose + D-galactose.  
 CC -1- CATALYTIC ACTIVITY: Glycosyl-N-acetylphingosine + H(2)O = a sugar +  
 CC N-acetylphingosine.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Brush border.  
 CC -1- TISSUE SPECIFICITY: Intestine.  
 CC -1- DOMAIN: THE SEQUENCE EXHIBITS 4 REGIONS (I-IV) OF INTERNAL  
 CC HOMOMOLOGY; THEREFORE LPH MIGHT HAVE EVOLVED BY TWO CYCLES OF  
 CC PARTIAL GENE DUPLICATION.  
 CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; X07995; CAA30802.1; -  
 CC HSSP; P26205; 1CBG.  
 DR HSSP: P26205; 1CBG.  
 DR InterPro: IPR001360; Glyco\_hydro\_1.  
 DR Pfam: PF00232; Glyco\_hydro\_1; 5.  
 DR PRINTS; PR00131; GLYHDRLASE1.  
 DR ProDom: PD000650; Glyco\_hydro\_1; 4.  
 DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 2.  
 DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F2\_2; 3.  
 KM Hydrolase; Glycosidase; Zymogen; Signal; Transmembrane; Repeat.  
 FT SIGNAL 1 19  
 FT PROPEP 20 866  
 FT CHAIN 867 1926  
 FT DOMAIN 20 1882  
 FT TRANSMEM 1883 1901  
 FT DOMAIN 1902 1926  
 FT DOMAIN 79 1800  
 FT REPEAT 79 172  
 FT REPEAT 360 845  
 FT REPEAT 881 1367  
 FT REPEAT 1375 1800  
 FT ACT\_SITE 1063 1063  
 FT ACT\_SITE 1271 1271  
 FT ACT\_SITE 1536 1536  
 FT ACT\_SITE 1747 1747  
 SQ SEQUENCE 1926 AA; 217847 MW; 2A21A7370D0CC7A CRC64;

Query Match 32.24; Score 852.5; DB 1; Length 1926;  
 Best Local Similarity 38.04; Pred. No. 7,4e-58;  
 Matches 184; Conservative 82; Mismatches 169; Indels 49; Gaps 13;

QY 8 PNDPWFATTAAYOIEGAVGEGSPSIMPYTCHESSSRNGANGVADACHYRHYDDFD 67  
 DB 381 PGGELWGVSTGAFVNEGMAEGGSPVWDQFGLKAAQ-QGATPEVASDSYKKMASDVA 439  
 QY 68 LITRYGAKARFSWGRIRPLGRLDPVNEEGIEFYSKIDALLRRGTPWVTLVWMDL 127  
 DB 440 LNLGGLAQVYKFSWSRIRPFMGSGSP-SPGCAVAYNKLIDSLDSHIEPMATLFWMDL 498  
 QY 128 PQAHLDRYGMVLNVEVOLDPERYARLCFERFGDRVONWITTXNPMIQAIYVATGSNAP 187  
 DB 499 PQAALDE-GGMON-ESVVDADFVDYAAFCFAFGNRLVMTFHEPVMVSAYAGCTGQHAP 556

QY 188 GRSSINKHSTEGNTATEPWLA-----GKAQIMSHARAVAAYSRDPRESQKQIGISNGDY 243  
 DB 557 GIS-----DPIGLASFOVAHLVVKANARFMTHYHNSHRQOQGRVGI VINSDW 603  
 QY 244 YEPWDSNEPRDKAERMEFHIGWPAFPIFLKKDYDESKKOLGER-----LPA 233  
 DB 604 AEPLESPERPEDLLAASERFLHFMIGWPAHPFVGDGDPATWKAQIQORNEQCPSPVAOLPE 663  
 QY 294 LTPADPAIINAGETDYGNNVYTSOFARHL--DGPVPEVDYGAINEHOENKQSPGEE 351  
 DB 664 FTTEKQQLK-GSADPLVGLSHYTSRLISKAPEDSCTPSYDTIGFSQHTD----PWPQ 717  
 QY 352 SGLAWPSCPDMPKRLARVYGLYGR---PIYITENGCCPCGGEENMTCEBAVNDPPRIYR 408  
 DB 718 TSSPWRIVPWPWGRIRLLQFVSLEYTGKVPYIYLAGNMPGISENL-----LDSLRVDY 772  
 QY 409 FDSHLDISKAITQDGVVYKGFAPMALLDNLEWSDGYPFGVYFTDY--TLTKRTPKS 466  
 DB 773 FNGYINEVLKAIKEDSDVDVARSYARSLMDGFEQAGYSQRFGLYHVNFNESKRTPRKS 832  
 QY 467 ALVL 470  
 DB 833 AFLV 836

RESULT 6  
 LPH\_RAT STANDARD; PRT; 1928 AA.  
 ID 002401; Q63712; Q63719;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Lactase-phlorizin hydrolase precursor (Lactase-glycosylceramidase)  
 DE [includes: Lactase (EC 3.2.1.108), Phlorizin hydrolase (EC 3.2.1.62)].  
 GN LCT OR LPH.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mistar; TISSUE=Intestine;  
 RX MEDLINE=91365258; Pubmed=1909681;  
 RA Duluc I., Boukamel R., Mantel N., Semenza G., Raul F., Freund J.-N.;  
 RT "Sequence of the precursor of intestinal lactase-phlorizin hydrolase  
 RT from fetal rat."  
 RL Gene 103:275-276(1991).  
 RN [2]  
 RP SEQUENCE OF 1-192 FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=93091244; Pubmed=1339333;  
 RA Boukamel R., Freund J.-N.;  
 RT "The rat LPH gene 5' region: comparative structure with the human  
 RT gene."  
 RL DNA Seq. 3:119-121(1992).  
 CC -1- FUNCTION: LPH splits lactose in the small intestine.  
 CC -1- CATALYTIC ACTIVITY: Lactose + H(2)O = D-glucose + D-galactose.  
 CC -1- CATALYTIC ACTIVITY: Glycosyl-N-acetylphingosine + H(2)O = a sugar +  
 CC N-acetylphingosine.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Brush border.  
 CC -1- TISSUE SPECIFICITY: Intestine.  
 CC -1- DOMAIN: THE SEQUENCE EXHIBITS 4 REGIONS (I-IV) OF INTERNAL  
 CC HOMOMOLOGY; THEREFORE LPH MIGHT HAVE EVOLVED BY TWO CYCLES OF  
 CC PARTIAL GENE DUPLICATION.  
 CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sb-sib.ch).

CC EMBL; X56748; CAA40070.1; -

DR EMBL; X56747; CAA40069.1; -

DR EMBL; L04635; AAA41539.1; -

DR PIR; J50610; J50610.

DR HSSP; P26205; 1CBG.

DR InterPro; IPR001360; Glyco\_hydro\_1.

DR Pfam; PF00232; Glyco\_hydro\_1; 5.

DR PRINTS; PR00131; GLYDRLAS1.

DR PRODOM; PD000650; Glyco\_hydro\_1; 4.

DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.

DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 2.

KW HydroLase; Glycosidase; Zymogen; Signal; Transmembrane; Repeat.

FT SIGNAL 1 19

FT PROPEP 20 867

FT CHAIN 868 1928

FT DOMAIN 22 1883

FT TRANSMEM 1884 1902

FT DOMAIN 1903 1928

FT DOMAIN 89 1842

FT REPEAT 89 175

FT REPEAT 365 849

FT REPEAT 886 1366

FT REPEAT 1371 1842

FT ACT\_SITE 1067 1274

FT ACT\_SITE 1274 1274

FT ACT\_SITE 1539 1539

FT ACT\_SITE 1750 1750

FT CONFLICT 7 7

FT CONFLICT 113 113

FT CONFLICT 207 207

FT SEQUENCE 1928 AA; 217266 MW; 56DDCAAC4ACAE85 CRC64;

Query Match 32.2%; Score 851.5; DB 1; Length 1928;

Best Local Similarity 40.5%; Pred. No. 8.9e-58;

Matches 193; Conservative 71; Mismatches 170; Indels 43; Gaps 13;

QY 8 PNDEWEGATAAVOTLEGAVKEGGRPSIMPTVYCHLEPERTGANGDVACDHVHYRDEPDFD 67

DB 1379 PKGFIWASAASAYVEGAMRADGKLSIMDTPFSRPLRIGDNDGDVACDSYHKAIDVY 1438

QY 68 LITRYGAKAYRPSLSWSRIIPLGRLDPVNEEGIEFYSKLIDALLRCIGTPWVTLVHMDL 127

DB 1439 ALQWGVSHYRFSIAMSRIIP-DGTTKFINAGLSYVRFIDALLAAGITPQVITVHMDL 1497

QY 128 POALHRRGGLNVEEVLDFERYARLCFERFGDRVQWITTKNPMIOAITGYATGSNAP 187

DB 1498 POALQD-VGWMEN-ETIVQREKAYADVLFQRLGDRVKVKTILNEPVIIAAGYGTGVSA 1555

QY 188 GRSSINKSTEGNTATEPWLAKQAQIMSHARAVAVSRDPSPQKQIGTISLNGDYVPEW 247

DB 1556 GISF-----RPTATYIAGHNLIKHAKAMHLYNVYARQCGTISITISDWEPR 1607

QY 248 DSNEDPRKEAERMEFHIGFANPIFKQDYPESMK-----KQLG-----ERLPALTPA 297

DB 1608 DPTREHREARASYQFWMGFAHPIFKNGDYPEVMKTRIDRSIGAGLANKSRLEPFTES 1667

QY 298 DPAIINAGETDPEYGNVYTSQFAHRLDGPVETTYLGIHGHQNKDQSPVGE-----SG 353

DB 1668 EKSRIR-GTFDFPGFNHTVLAIVL-----DYPAFSSFDADRGVAIASDSSWPVSG 1719

QY 354 LAWLRSCPDMPFRKLHARVGLYGR-PIYITENGCPCEENMTCEEAIVNDPFRIRYFDSH 412

DB 1720 SFWLKVTPFGRRRIIMNLKEEYNNPPIYITENGSRREPEL-----NDIDRIYLRSY 1773

QY 413 LDISKAITQDGVVVKGFAMALLDNIEMSDGYGRGVTFTDYT--TLKRTPKKSA 467

DB 1774 INEALKAV-HDKVLDRGYTVWSIMDNFEMATGFAERFVHVNVSNDPSPRIPRASA 1829

ID MYO\_ARATH STANDARD; PRT; 541 AA.

AC P37702; O8H7H2; Q93231; Q940N8;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Myrosinase precursor (EC 3.2.1.147) (sinigrinase) (thioglucosidase).

GN TCGL OR AT5G26000 OR T1N24.7.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

QY 1

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

EX MEDLINE=94302158; PubMed=8029343;

RA Chadehwan S., Bishop J., Thangstad O.P., Bones A.M.,

RA Mitchell-Olds T., Bradley D.;

RT "Arabidopsis cDNA sequence encoding myrosinase.";

RL Plant Physiol. 103:671-671 (1993).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=leaf;

RX MEDLINE=95284348; PubMed=7766881;

RA Xue J., Joergensen M., Philgren U., Raek L.;

RT "The myrosinase gene family in Arabidopsis thaliana: gene

RL organization, expression and evolution.";

PL Plant Mol. Biol. 27:911-922 (1995).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016721; PubMed=11130714;

RA Tabata S., Kaneo T., Nakamura Y., Korani H., Kato T., Asamizu E.,

RA Miyajima N., Sasekoto S., Kimura T., Hosouchi T., Kawashima K.,

RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,

RA Nakazaki N., Natsu K., Okumura S., Shimo S., Takeuchi C., Wada T.,

RA Wakabayashi A., Yamada M., Yasuda M., Sato S., de la Bastide M.,

RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,

RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,

RA Stonking T., Pegin K., Spieth J., Sekhon M., Armstrong J., Becker M.,

RA Belter E., Cordum H., Cordes M., Courtney L., Courtney L., Danne M.,

RA Du H., Edwards J., Fryman J., Haakenen B., Lamar E., Latreille P.,

RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,

RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,

RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,

RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,

RA Mattiensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,

RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,

RA Vanlaere K.-D., Terry N., Hartley N., Bent E., Johnson S.,

RA Langham S.-A., McCullagh B., Roben J., Gyomai B., Zimmermann W.,

RA Ramsperger U., Medler H., Balke K., Medler E., Peters S.,

RA van Staveren M., Dirke W., Moollman P., Klein lankhorst R.,

RA Weltzner T., Bothe G., Rose M., Hauf J., Bernier S., Hempel S.,

RA Feldpuscher W., Lamberth S., Villarroel R., Gelen S., Ardiles W.,

RA Berts O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,

RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.;

RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis

RL thaliana.";

RL Nature 408:823-826 (2000).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=22954850; PubMed=14593172;

RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,

RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,

RA Katlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,

RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,

RA Chan M.M., Tang C.C., Onders C.S., Deng J.M., Akiyama K., Ansari Y.,

RA Chao K., Bohn J., Banno F., Bowser L., Brooks S.Y., Carrinci P.,

RA Chao Y., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,

RA Hayashizaki Y., Johnson-Hopson C., Hsuan Y.W., Iida K., Karnes M.,

RA Khan S., Koesema E., Ishida J., Jang P.X., Jones T., Kawai J.,

RA Kamiya A., Meyers C., Nakajima M., Natsuka M., Seki M., Sekurai T.,

RESULT 7

MIRO\_ARATH

RA	Sacou M., Tamme R., Vayenberg M., Wallender E.K., Wong C.J., Yamamura Y.,
RA	Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.
RT	"Empirical analysis of transcriptional activity in the Arabidopsis
RT	genome.";
RL	Science 302:842-846(2003).
RN	[5]
RN	SEQUENCE OF 1-226 FROM N.A.
RP	Stracke R., Palme X.;
RA	"Signal peptide selection derived cDNAs from Arabidopsis thaliana
RT	leaves and guard cells."
RL	Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN	[6]
RC	SEQUENCE OF 192-250 FROM N.A.
RP	STRAIN=cv. C24; TISSUE=flower buds;
RA	Mache R., Quigley F., Thomas F., Yu D.Y.;
RL	Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
CC	-1- FUNCTION: Degradation of glucosinolates (glucose residue linked by
CC	a thioglucoside bound to an amino acid derivative) to glucose,
CC	sulfate and any of the products: thiocyanates, isothiocyanates,
CC	nitriles, epithionitriles or oxazolidine-2-thione.
CC	-1- CATALYTIC ACTIVITY: H(2)O + a thioglucoside = a sugar + a thiol.
CC	-1- SUBUNIT: Homodimer (By similarity).
CC	-1- SUBCELLULAR LOCATION: Vacuolar (By similarity).
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event-Alternative splicing: Named isoforms=1;
CC	Comment=A number of isoforms are produced. According to EST
CC	sequences;
CC	Name=1;
CC	IsoId=P37702-1; Sequence=Displayed;
CC	-1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> )
CC	or send an email to <a href="mailto:license@isb-sdb.ch">license@isb-sdb.ch</a> ). -----
DR	EMBL; L11454; AAC18869.1; -
DR	EMBL; X79194; CAAS5786.1; -
DR	EMBL; AF149413; AAD40143.1; -
DR	EMBL; AY045681; AAK74039.1; -
DR	EMBL; AY054237; AAL06896.1; -
DR	EMBL; AY058182; AAL25596.1; -
DR	EMBL; AY090382; AAL91284.1; -
DR	EMBL; AF083677; AAN60236.1; -
DR	EMBL; Z18232; CAAT9143.1; -
DR	PIR; S56653; S56653.
DR	HSSP; P26205; ICBG.
DR	SWISS-2DPAGE; P37702; ARATH.
DR	InterPro; IPRO01360; Glyco_hydro_1.
DR	Pfam; PF00232; Glyco_hydro_1; 1.
DR	PRINTS; PRO0131; GLHYDRLASE1.
DR	ProdDom; PD000650; Glyco_hydro_1; 1.
DR	PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR	PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW	Hydrolase; Glycosidase; Signal; Glycoprotein; Alternative splicing.
FT	SIGNAL
FT	1                19
FT	CHAIN
FT	20              541
FT	ACT SITE
FT	420          420
FT	CARBOHYD
FT	33           33
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	108          108
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	175          175
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	236          236
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	356          356
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	379          379
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	493          493
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CARBOHYD
FT	512          512
FT	N-LINKED (GLCNAC . . . ) (POTENTIAL)
FT	CONFLICT
FT	385          385
FT	P -> H (IN REF. 4; AA106896).
FT	CONFLICT
FT	426          426
FT	P -> A (IN REF. 4; AA125596).
FO	SEQUENCE
FO	541 AA: 61132 MW: 3736B735DE7A5BD1 CRC64;

Query Match 32.1%; Score 849.5; DB 1; Length 541;  
Best Local Similarity 39.5%; Pred. No. 2,4e-58;  
Matches 185; Conservative 87; Mismatches 157; Indels 39; Gaps 14;

OY 11 FEMGPATAAAYOIEGAVKEGGRGPSIMDTYCHLEBSRTNGA--NGDVA CDHYHRDEDFD 67  
DB 47 FIFGVASSAYOVEGG---RGRGLNVWDSFTHRPPEK-GGADLGDGTTCOSYTLMQID 102  
OY 68 LUTTYGAKARFSLSNRRIITPLGRLDPPVNEEGEFISKILDALRRGITPWTLYWMDL 127  
DB 103 VMDELINSTGVAFSIAWSRLLPKGRSRGNVPGAIKYNNGLIDGLVANMTPFTVLFFHWDL 162  
OY 128 PQALHDRGGLANVEVOLDEPERARLCFEFRFGVRGNVMITIXNPMIQAIVTGSNAP 187  
DB 163 PQTLYDEVNGFLNKTIYD-DPKYADJADCFELFDGRVKMWITINDLYVPTPRIGVALGTDAP 221  
OY 188 GRSS--INKSTEGNTATPEWLHGKAQIMSHARA VAAYSRDPRSPKOQIGISLNGDYE 245  
DB 222 GRGSPKIDVRCPGNSSSTEPIYVHNQLLMAAAVADVTRTKYKDQDGGMIGPVITRMFL 281  
OY 246 PWDSNEPDDEAEARRNEFHIGWANIIFLKQYPESMKOLGERLLPALPADPAIIAAG 305  
DB 282 PFDSHO-EBSKATERRAKIFPHGWMTGLETEKG-YPDIMREVGVRLEPFESTEAIVK-G 338  
OY 306 ETDYGNMYNYSGPARHLDDGVPTDYLGAIHER---QENKOSPVE-----ES 352  
DB 339 SYDLPLGNLYVTQYAONNQTTIPSDVHTALMDSRTTLTSKNATHABCPPEMAASYPYK 398  
OY 353 GLAMLRSCPDMPFRGHRLRVGLYKP-IYITENCPCPGEENMTCEBAVNDPFRIRYDFS 411  
DB 399 GIYYVV---DYFKT-----TYDDPLIYLTENGFSFRGDED-F-EKATYDKRIDVLCS 446  
OY 412 HLDSISAKITQDVGVKGYFAALLDNLEMSDYGPRFGVTFDYTL 459  
DB 447 HLCFLSKVIKEKNVNKGYFAMSIGDYVEFCNGFTVFGLSYVDPAFI 494

RESULT 8  
MYR3\_SINAL  
ID MYR3 SIGNAL STANDARD; PRF; 544 AA.  
AC P23092;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Myrosinase MB3 precursor (EC 3.2.1.147) (Sinigrinase)  
DE (Thioglucosidase).  
OS Sinapis alba (White mustard) (Brassicaceae; Brassicaceae; Sinapis).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosoids II; Brassicales; Brassicaceae; Sinapis.  
OX NCBI\_TaxID=3728;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=cv. Maxi; TISSUE=Seed.  
RX MEDLINE=92119259; PubMed=1731996;  
RA Xue J.; Lemmen M.; Falk A.; Raak L.;  
RT "The glucosinolate-degrading enzyme myrosinase in Brassicaceae is encoded by a gene family.";  
RU Plant Mol. Biol. 18:387-398(1992).  
CC - FUNCTION: Degradation of glucosinolates (glucose residue linked by a thioglucoside bound to an amino acid derivative) to glucose, sulfate and any of the products: thiocyanates, isothiocyanates, nitriles, epithionitriles or oxazolidine-2-thiones.  
CC - CATALYTIC ACTIVITY: H(2)O + a thioglucoside = a sugar + a thiol.  
CC - SUBUNIT: Homodimer.  
CC - SUBCELLULAR LOCATION: Vacuolar.  
CC - TISSUE SPECIFICITY: In vacuoles called myrosin grains of a certain class of cells, myrosin cells, distributed in the cotyledons and the axis of the embryo as well as in different organs of the growing plant.  
CC - SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
CC This SWISS-PROT entry is copyright It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

DR EMBL: X59879; CAA42534.1; -.  
 DR PIR: S19149; S19149.  
 DR HSP: P26205; ICBG.  
 DR InterPro: IPR001360; Glyco\_hydro\_1.  
 DR Pfam: PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS: PR00131; GLHYDRLASE1.  
 DR ProDom: PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE: PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
 DR PROSITE: PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 DR HydroLase; Glycosidase; Multigene family; Glycoprotein; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 544 MYROSINASE MB3.  
 FT ACT SITE 426 446 NUCLEOPHILE (BY SIMILARITY).  
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 240 240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 544 AA; 62050 MW; 3971D304CF75C28 CRC64;

Query March 32.0%; Score 845; DB 1; Length 544;  
 Best Local Similarity 39.3%; Pred. No. 5.3e-58;  
 Matches 190; Conservative 86; Mismatches 158; Indels 50; Gaps 15;

QY 10 DFEWGFATTAAYQIGAVNVEGGGSGPSIMDTYCHLESRRING--ANGDVACHYHREDFD 67  
 DB 48 DFIGVASSAQIEG---RGRGVNVMDGFSHRPEKSGSLKNGDTSCEBYTMKKDVE 104  
 QY 68 LITRYGAKAYRFSLSWSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHMDL 127  
 DB 105 IMGELNATGYRFSFAMSRIVPKKVRSGVDQAGLDYHNLDALLEKNITPFTVLLFHWDL 164  
 QY 128 PQLALHDYGGWLNVEEVLDFERYARLCFERFGDRVQNMITINXPMIOAIYGYATGSNAP 187  
 DB 165 PQLQDEYEGFLDRQIIO-DFKDYADLCFEKFGGVKVMITINOLYTVPTRGVALGTDAP 223  
 QY 188 GRSSINKHST---GNTATEPMLAKGAQIMSHARAANVAVSRDRPSQKQIGISLNGDY 243  
 DB 224 GRGCPYDTKRCRCGSGNSTEPIYVANHQLAHAIYDLVYTN-AFONGKIGVMTIRW 282  
 QY 244 YEPWDSNEPRDKAERRMERFHIQFANPIFLKKDYPSMKKQIGERLPALTPADFAILN 303  
 DB 283 FLPYDESDPACIEAERNQCFHGMWMEPL-TKGRYPDIMEQIVGSRLPFTTEAE-AELV 340  
 QY 304 AGEIDFYGMNYTTSQFAHLDGPVP-ET-----DYLGAIHHEQNKD 344  
 DB 341 AGSYDFGLNLYVYQAKPKNPYPSETHIALMDAGVDLTNNRSGEYGGVFAEDANSY 400  
 QY 345 GSPVGEESGLMLRSCPMFRKHLARVYGLYKGP-IYTTENGCPGGEENNTCEEAANDP 403  
 DB 401 YYPKG-----IYVMDYFKTK-----YNNPLIYITTEGISTPSES-RC-EALADY 444  
 QY 404 FRIYFSDHDSISKAITQDGVVVKGFAMALLDNLEWSDGYGPPFGVTFTDYTTL-KRT 462  
 DB 445 KRINVLCSHLCFLRKVIREKGVNIRGFYAMALGDNVEFCFGKFTVRFGLSYVNWMDLDRN 504  
 QY 463 PKKS 466  
 DB 505 LKES 508

RESULT 9  
 BGLA\_THEME STANDARD; PRT; 446 AA.  
 AC Q08638;  
 DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, last sequence update)  
 DT 15-MAR-2004 (Rel. 43, last annotation update)  
 DE Beta-glucosidase A (EC 3.2.1.21) (Gentiobiose) (Cellulobiose) (Beta-D-  
 DE glucoside glucosylhydrolase).  
 GN BGLA.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.

OX NCBI\_Taxid=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;  
 RX MEDLINE=94104595; PubMed=8277941;  
 RA Liebl W., Gabelsberger J., Schleifer K.H.;  
 RT "Comparative amino acid sequence analysis of Thermotoga maritima  
 RT beta-glucosidase (Bgla) deduced from the nucleotide sequence of the  
 RT gene indicates distant relationship between beta-glucosidases of the  
 RT BGA family and other families of beta-1,4-glycosyl hydrolases.";   
 RL Mol. Gen. Genet. 242:111-115(1994).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
 CC -1- glucose residues with release of beta-D-glucose.  
 CC -1- PATHWAY: Cellulose degradation.  
 CC -1- MISCELLANEOUS: THE DNA CODING FOR THIS PROTEIN IS NOT FOUND IN THE  
 CC COMPLETE GENOME OF T.MARITIMA. IT COULD HAVE ORIGINATED FROM  
 CC ANOTHER BACTERIAL SPECIES.  
 CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

DR EMBL: X74163; CAA52276.1; -.  
 DR PIR: S41561; S34570.  
 DR HSP: P22073; ITR1.  
 DR InterPro: IPR001360; Glyco\_hydro\_1.  
 DR Pfam: PF00232; Glyco\_hydro\_1; 1.  
 DR PRINTS: PR00131; GLHYDRLASE1.  
 DR ProDom: PD000650; Glyco\_hydro\_1; 1.  
 DR PROSITE: PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
 DR PROSITE: PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
 KW HydroLase; Glycosidase; Cellulose degradation.  
 FT ACT SITE 166 166 PROTON DONOR (POTENTIAL).  
 FT ACT SITE 351 351 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 446 AA; 51548 MW; 2B5B0672B8F4C01 CRC64;

Query March 31.7%; Score 839; DB 1; Length 446;  
 Best Local Similarity 41.1%; Pred. No. 1.2e-57;  
 Matches 193; Conservative 76; Mismatches 145; Indels 56; Gaps 19;

QY 8 PNDWGFATTAAYQIGAVNVEGGGSGPSIMDTYCHLESRRINGANGDVACHYHREYDDFD 67  
 DB 7 PEGFLMGVATAASAYQIESPLADGAGMSIMHTFSTPNVNGDGDVACHYNNWKEDIE 66  
 QY 68 LITRYGAKAYRFSLSWSRIIPLG-GRLDPVNEEGIEFYSKLIDALLRGITPWTLYHMD 126  
 DB 67 IIEKLGAKAYRFSLSWSRIIPLEGTR--VNQKGLDPTNNKIIDLLEKGIPTFTIYHMD 123  
 QY 127 LPOLALHDYGGWLNVEEVLDFERYARLCFERFGDRVQNMITINXPMIOAIYGYATGSNA 186  
 DB 124 LPPALQK-GGMAN-REIADWFAYSRVLPENFGDRVQNMITINPPWVAIVGHLVGVHA 181  
 QY 187 PGRSSINKHSTEGNTATEPMLAGKA--QMSHARAANVAVSRDRPSQK-QQIGISLNGD 242  
 DB 182 PGMKDI-----YVAFVAVNLLRAHRAVAV-----FRETQKQKGIIVNNG 224  
 QY 243 YIEPWSNEPRDKAERRMERF-HIGFANPIFLKKDYPSMKKQIGERLPALTPADFAI 301  
 DB 225 YIEP-ASEKEDIRAVAFMIOFNYPFLFNPYI-RGIYPELVLEFAEYLPENYKDDMS 282  
 QY 302 LNAGETDFGMNYTTSQFAHLDGPVPETDYLGAIHHEQNKDQSPVGE--SGLAMLR 358









Db 124 POALQDQ-GCM--GSRITIDAFAYELMFKELGKIKOMITENPWCMAFLSNYLGVHA 180  
 Qy 187 PGS-----SINKHSTEGTATPEMLAGKAQIMSHARAVASRDPRPOKQIGISLN 240  
 Db 181 PGKNDQLAIDVSHH-----LVAHGRAVTLF-RELGIS--GEIGIAPN 221  
 Qy 241 GDYVPEPDSNEPDKAEARMEFHIGFANPIFLKKDYPSM---KKQGERLPALTPA 297  
 Db 222 TSMAVPRRTK-EDMECLARNGMSGWYLDPIYF-GEYKFMIDWENIGYK-PRIVDG 278  
 Qy 298 DPAIINAGETPDYGMNYYTSQFARHLDPVPEYDYLGAIEHODENKDGSPVGEESGLAWL 357  
 Db 279 DMELIIG-PIDFICINYYTSSMKNYNPGEA-----GCMLESAISMGAP---KTIDIGM- 327  
 Qy 358 RSCPEMRKRLARYGL-----YGRP-IYITENGCPGCEEMTCEAVNDPRIRYF 409  
 Db 328 ----EYAEGL---YDLRYTADRYGNPTLYTENGA-C-YNDGLSLDGRHIDRRIDYL 378  
 Qy 410 DSHLDSISKAITODGVVVKGYFAMALLDNLEWSDGPRFGVTFTDTTLKRTPKKS 466  
 Db 379 AMHLIQASRAI-EDGINKGYMEWMLDNFEMAGYCMRGVLHVVDYDVLVTRPKDS 434

## RESULT 13

BGLB\_PABPO

ID\_BGLB\_PABPO STANDARD; PRT; 448 AA.

AC P2505;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Beta-glucosidase B (EC 3.2.1.21) (Gentloblase) (Cellloblase) (Beta-D-glucoside glucosylhydrolase) (Amygdalase).

GN BGLB. *Bacillus polymyxa* (*Bacillus polymyxa*).OS *Bacillus polymyxa* (*Bacillus polymyxa*).

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI\_TaxID=1406;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91071603; Pubmed=2123813;

RA Gonzalez-Candelas L., Ramon D., Polaina J.;

RT "Sequences and homology analysis of two genes encoding

beta-glucosidases from *Bacillus polymyxa*."

RL Gene 95:31-38(1990).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-glucose residues with release of beta-D-glucose.

CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).

CC EMBL: M60211; AAA23264.1; -

DR PIR; J00038; J00038.

DR HSSP; P22073; ITR1.

DR InterPro; IPR001360; Glyco\_hydro\_1.

DR Pfam; PF00232; Glyco\_hydro\_1; 1.

DR PRINTS; PR00131; GMYDRLASE1.

DR Prodom; PD000650; Glyco\_hydro\_1; 1.

DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.

DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.

KW Hydrolyase; Glycosidase; Cellulose degradation.

FT ACT\_SITE 167 PROTON DONOR (POTENTIAL).

FT ACT\_SITE 167 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 448 AA; 51573 MW; 628F44507C21EPIF CRC64;

Query Match 29.9%; Score 791; DB 1; Length 448;

Best Local Similarity 38.1%; Pred. No. 6.4e-54;

Matches 185; Conservative 61; Mismatches 172; Indels 68; Gaps 15;

Qy 8 PNDPEWGFATAAYQIEGAVEGGRSPINDTYCHLEPSRTNGANGVACDHYRVEDFD 67  
 Db 9 PATFMGTSTSSYSQISGGDEGRTPSINDTFQIQGVKIGCGQVACDHPFHFKEDVQ 68  
 Qy 68 LTRKGAAYRSLSSRIIPLEGRLDPVNEBIEFYSKILDLLRGRITPWTLLYMDL 127  
 Db 69 LMKQFLHYHRFSVAMPRIIMPAG---IINEBGLFYEHLDEIEAGLPMULTLYHMDL 125  
 Qy 128 POALHRYGGMVAVEVQDLEFERYARLCPEFGDRVQNNITTXPMIOATYVATGSMA 187  
 Db 126 PQMTEDB-GGWTORETIQ-HFKTYASVIMDRFEBRINMNTINBPYCASTLGIGTGEHAP 183  
 Qy 188 GRSISINKHSTEGTATEPMLAGKAQIMSHARAVASRDPRPOKQIGISLNGDYEPW 247  
 Db 184 GHENMR-----EAFIAHILMCHGIAANLHK---EKLGTGIGITLMEHAYDA- 229  
 Qy 248 DSNPEPDKAEARMEFHIGFANPIFLKKDYPSMKQIGERLPAL---TPADFAILNA 304  
 Db 230 ASRRPEVAAAIRDDGINRWFAEPLFNCR-YPEDWVWYGYLNGLDVQPDMDLIQ 288  
 Qy 305 GEDTFYGMNYYTSQFAR-----HLDGVPETDYLGAIHEHODENKDGSPVGE 351  
 Db 289 -PGDFGINYYTSITRSTNDASLQVEQVHMEEPY--TDMGEIH----- 331  
 Qy 352 SGLAWLRSCPEMRKRLARYGLYGR--PIYITENGCPGCEEMTCEAVNDPRIRYF 409  
 Db 332 -----PESTFKULTREKDFSGKGLPILITENGAARDE---LVNGQIEDTGRHGYI 379  
 Qy 410 DSHLDSISKAITODGVVVKGYFAMALLDNLEWSDGPRFGVTFTDTTLKRTPKSALV 469  
 Db 380 EEHLKACHREI-BEGQLKGYFWMSFLDNFEMAGYSKRGIVHINVEOTERTPKOSALW 438  
 Qy 470 LKDMFA 475  
 Db 439 FKQWMA 444

## RESULT 14

BGLT\_TRIRP

ID\_BGLT\_TRIRP STANDARD; PRT; 425 AA.

AC P26205;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cyanogenic beta-glucosidase precursor (EC 3.2.1.21) (Linamarase) (Fragment).

GN *Linum catharticum* (Creeping white clover).

OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Trifolium.

OX NCBI\_TaxID=3899;

RN [1]

RP SEQUENCE FROM N.A. AND SEQUENCE OF 12-25 AND 125-147.

RC STRAIN=SI00 (EG); TISSUE=leaf;

RX MEDLINE=91322517; Pubmed=1907511;

RA Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;

RT "Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (*Trifolium repens* L.).";

RL Plant Mol. Biol. 17:209-219(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).

RX BARRETT T., Suresh C.G., Tolley S.P., Dodson E.J., Hughes M.A.;

RT "The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase.";

RL Structure 3:951-960(1995).

CC -1- FUNCTION: Hydrolyses cyanoglucosides, contributing to the release

CC of hydrocyanic acid, which functions as a defense mechanism

CC against small predators, when the leaf tissue is damaged.

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-glucose residues with release of beta-D-glucose.

CC -1- PATHWAY: Cyanogenesis.

CC -1- SUBUNIT: Homodimer.  
CC -1- TISSUE SPECIFICITY: Leaves.  
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X56733; CAA40057.1; -  
DR PIR; S16580; GLJY14.  
DR PDB; 1CBG; 15-OCT-95.  
DR InterPro; IPR001360; Glyco\_hydro\_1.  
DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
DR PRINTS; PR00131; GLYDRLASE1.  
DR PRODOM; PD000650; Glyco\_hydro\_1; 1.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1\_1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F1\_2; 1.  
KW Hydrolase; Glycosidase; Glycoprotein; Signal; 3D-structure.  
FT NON TER 1  
FT SIGNAL 1  
FT CHAIN 1  
FT ACT\_SITE 12 425  
FT ACT\_SITE 194 425  
FT ACT\_SITE 408 408  
FT DISULFID 213 221  
FT CARBOHYD 220 220  
FT CARBOHYD 412 412  
FT TURN 16 17  
FT HELIX 22 24  
FT HELIX 27 29  
FT TURN 32 33  
FT STRAND 35 39  
FT HELIX 42 45  
FT TURN 52 53  
FT STRAND 57 57  
FT HELIX 58 65  
FT HELIX 67 69  
FT TURN 71 72  
FT TURN 79 80  
FT HELIX 82 95  
FT TURN 96 97  
FT STRAND 100 104  
FT HELIX 107 110  
FT TURN 112 113  
FT HELIX 116 118  
FT HELIX 122 137  
FT TURN 138 139  
FT STRAND 141 147  
FT STRAND 152 152  
FT HELIX 153 159  
FT HELIX 161 163  
FT TURN 165 166  
FT HELIX 167 182  
FT TURN 183 185  
FT STRAND 188 193  
FT HELIX 195 203  
FT TURN 204 204  
FT TURN 209 210  
FT HELIX 215 217  
FT TURN 226 228  
FT HELIX 252 251  
FT TURN 252 252  
FT HELIX 253 256  
FT STRAND 259 265  
FT STRAND 268 271  
FT HELIX 276 289  
FT TURN 290 290  
FT HELIX 291 299  
FT HELIX 304 310  
FT HELIX 311 313

FT HELIX 319 325  
FT TURN 326 327  
FT STRAND 331 335  
FT STRAND 339 344  
FT TURN 349 350  
FT HELIX 355 358  
FT TURN 359 359  
FT STRAND 361 364  
FT STRAND 366 367  
FT TURN 368 369  
FT STRAND 370 371  
FT STRAND 375 375  
FT TURN 379 380  
FT STRAND 382 382  
FT TURN 385 385  
FT HELIX 386 398  
FT TURN 399 400  
FT STRAND 404 408  
FT STRAND 413 414  
FT TURN 417 418  
FT HELIX 421 425  
SQ SEQUENCE 425 AA; 48324 MW; 20B983B65C47A678 CRC64;

Query Match 29.9%; Score 790; DB 1; Length 425;  
Best Local Similarity 43.9%; Pred. No. 7.1e-54;

Matches 169; Conservative 63; Mismatches 135; Indels 18; Gaps 10;

QY 11 FEMGATAVAQIEGAVKEGGRGSPISWDYCHLEPSR-TNGANGDVACDHVHRDEDFLL 69  
Db 34 FVEGTASAPAFQYGAFAFEDGKPSIWDTFHKKYPEKIKORTGNDVAIDYHRYKEDIGIM 93  
QY 70 TKYGAAYRPSLSMSRIIPLGRLDPVMEGIEFYKLIIDALIRGITWVTLHYMDLPQ 129  
Db 94 KDMNLDAYRPSISWPVLPRGKLSGVNREGINYNLLINEVLANGMOFVYTLFHWDPQ 153  
QY 130 ALHDIRYGMVINEVEVOLDFERVYARLCFERFGDRVQWMITINXPMQAIYVATGSNAPGR 189  
Db 154 ALDEYRGFLGRNIVD-DFRDYAEI CFKFGDRVGHMITLNEPKGVSMAAYVGFAPGR 212  
QY 190 SS--INKHTEGNTATEPWLAKQAQIMSHARAVAVYSRDFPSQGIQGISLNGDYERW 247  
Db 213 CSDWLKNCTGGSGSEPYLAHYQLAHMAAARLYKTYQASQNGIIGITLVSHWFEF- 271  
QY 248 DSNRPDRKEAERBERFHIQWFPANPIFLKQVPSMKQLGRRLPALTPADPAIINAGET 307  
Db 272 ASKEKADVDAAKRGIDPMLGMFMHP-TKGRYPESMRVLYRKLPSTESKEL-TGSF 329  
QY 308 DRYGMNYYTSQPARHLD----GPVETDYI-GAIHEQENKDGSPVGEESGLAWLRSCP 361  
Db 330 DFLGLNYYSSYYAAKAPRIPARPAIQTDSLINATFEH---NGKPLGPMMASSWLCIYP 385  
QY 362 DMFRKHLARVYGLYGP-ITYTENG 385  
Db 386 QGIRKLLLYVKNYNNPVYITENG 410

RESULT 15  
BGLA\_PAEPO  
ID BGLA\_PAEPO STANDARD; PRT; 448 AA.  
AC P22073;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta-glucosidase A (EC 3.2.1.21) (Gentliobase) (Celllobase) (Beta-D-  
DE glucoside glucosylase) (Amygdalase) (BGA).  
GN BGLA.  
OS Paenibacillus polymyxa (Bacillus polymyxa).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxId=1406;  
RP [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91071603; PubMed=2123813;  
RA Gonzalez-Candelas L., Ramon D., Polaina J.;

RT "Sequences and homology analysis of two genes encoding  
RT beta-glucosidases from *Bacillus polymyxa*.";  
RL Gene 95:31-38(1990).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
RC STRAIN=ATCC 842;  
RX MEDLINE=98139963; PubMed=9466926;  
RA Sanz-Aparicio J., Hermoso J.A., Martinez-Ripoll M., Lequerica J.L.,  
RA Polaina J.;  
RT "Crystal structure of beta-glucosidase A from *Bacillus polymyxa*:  
RT insights into the catalytic activity in family 1 glycosyl  
RT hydrolases.";  
RL J. Mol. Biol. 275:491-502(1998).  
CC -1- FUNCTION: BGLA IS INTRACELLULAR AND CLEAVES CELLOBIOSE PROBABLY  
CC THROUGH INORGANIC PHOSPHATE MEDIATED HYDROLYSIS.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-  
CC glucose residues with release of beta-D-glucose.  
CC -1- SUBUNIT: Homooctamer.  
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; J06037; AAA22263.1; -;  
DR PIR; J06037; J06037.  
DR PDB; 1BGA; 15-APR-98.  
DR PDB; 1BGG; 27-MAY-98.  
DR PDB; 1TR1; 20-APR-99.  
DR PDB; 1E41; 05-JUL-01.  
DR InterPro; IPR001360; Glyco\_hydro\_1.  
DR Pfam; PF00232; Glyco\_hydro\_1; 1.  
DR PRINTS; PR00131; GLYHYDRLASE1.  
DR PRODOM; PD000650; Glyco\_hydro\_1; 1.  
DR PROSITE; PS00572; GLYCOSYL\_HYDROL\_F1; 1.  
DR PROSITE; PS00653; GLYCOSYL\_HYDROL\_F2; 1.  
KW Hydrolase; Glycosidase; Cellulose degradation; 3D-structure.  
FT ACT SITE 166 166 PROTON DONOR (POTENTIAL).  
FT ACT SITE 352 352 NUCLEOPHILE (BY SIMILARITY).  
FT STRAND 3 5  
FT TURN 8 9  
FT STRAND 11 15  
FT HELIX 18 21  
FT TURN 25 26  
FT HELIX 27 29  
FT STRAND 33 33  
FT HELIX 34 39  
FT TURN 40 40  
FT TURN 42 44  
FT TURN 46 48  
FT HELIX 54 55  
FT TURN 57 71  
FT HELIX 72 72  
FT TURN 75 79  
FT STRAND 82 85  
FT HELIX 87 88  
FT TURN 95 110  
FT HELIX 111 112  
FT TURN 114 120  
FT STRAND 125 125  
FT STRAND 126 130  
FT HELIX 131 132  
FT TURN 133 135  
FT HELIX 137 137  
FT TURN 138 153  
FT HELIX 154 154  
FT TURN 155 157  
FT HELIX 160 165  
FT STRAND 167 175

FT TURN 176 176  
FT TURN 181 182  
FT HELIX 187 210  
FT TURN 211 212  
FT STRAND 216 220  
FT STRAND 222 222  
FT STRAND 226 228  
FT HELIX 233 246  
FT TURN 247 247  
FT HELIX 248 256  
FT HELIX 261 269  
FT TURN 270 271  
FT TURN 278 279  
FT HELIX 280 284  
FT TURN 285 285  
FT STRAND 290 294  
FT STRAND 299 303  
FT TURN 305 306  
FT TURN 308 311  
FT STRAND 312 314  
FT STRAND 321 321  
FT TURN 323 324  
FT STRAND 327 327  
FT TURN 330 330  
FT HELIX 331 340  
FT HELIX 341 343  
FT STRAND 348 352  
FT STRAND 362 362  
FT TURN 363 364  
FT STRAND 365 365  
FT HELIX 369 387  
FT TURN 388 389  
FT STRAND 392 397  
FT STRAND 401 401  
FT HELIX 406 411  
FT STRAND 417 419  
FT TURN 421 423  
FT STRAND 426 428  
FT HELIX 430 441  
FT TURN 442 442  
FT STRAND 443 445  
SQ SEQUENCE 448 AA; 51649 MW; D971D2E61B6627C1 CRC64;  
Query Match 29.8%; Score 789; DB 1; Length 448;  
Best Local Similarity 39.8%; Pred. No. 9.1e-54;  
Matches 184; Conservative 65; Mismatches 173; Indels 40; Gaps 18;  
QY 8 PNDFEWGFATTAAYQIEGAVKEGGGSPSIWDYCHLEPSRTNGANGDVACDHYHRDDEFD 67  
DB 7 PDDFMWGTATAAYQIEGAVQEDGRGLSINDTFATHTGKVFNGDNGNVACDSYHREEDIR 66  
QY 68 LTTKGAKAYRFLSRSRIIPLAGRLDPVNEBSIEBYSKIDALLRGITPWTLYIHWDL 127  
DB 67 LMKELGIRTYRFSVSPRIFFPND--GEVNOEGLDYHRVLDLNDNGIEPCTLYHWDL 124  
QY 128 PQLAHDYRGWMLNVEEVOLEFERYARLCFERFGDRVQNNMTTIXPMIOIYVATGSNAP 187  
DB 125 PQLAODA-GGNGKRRITOA-FVQFAETMFRREFGKIQHMTTFNEPNCIAFLSNMLGVHAP 182  
QY 188 GRSSINKHSTEGNTATEPMLAGKAQIMSHARAVAYSRDPR-SQKQIGISLNGDYEP 246  
DB 183 GLTNL-----QTAID---VGHHLVAVGLSV---RRRELGTSGQIGIAPNVSWAVP 228  
QY 247 WDSNEPRDKEAARRRNEPHIGWPAHYIFLKKDYPSMKQOLGRRLPALTLPADPAIINAGE 306  
DB 229 YSTSE-EDKAAACARTISLSHDFLOPIY-QGSYPOFLVWMAEQGATVPIDQDMOIGI 286  
QY 307 -TDYGANVYTSQFAHLDGPVETDYLGAIHGHQENKQSPVGESEGLAWLSCPFDMFR 365  
DB 287 PDMIGINYSVNF---NPEAGFL-----OSEINNGLPV---TDIGMPVESGLVE 335  
QY 366 -KHLARVYGLYGRPIYITENGCPGPEENMTCEAVNDPFRIRYFDSHLSISKAITODG 424

Mon Mar 29 09:02:57 2004

us-10-026-140-2.rsp

Page 14

Db 336 VLHYLQKYG--NIDIYITNGA-CINDE--VNGKGVQDDRISIMQGHVQVHRTI-HDQ 389

Qy 425 VVKGYPAMALLNLEMSDGYGRGVPTDITTLKRTPKKS 466

Db 390 LHYGVMAWGLDNFEMARGYNNRFGMLHVDPTQVTPPKES 431

Search completed: March 25, 2004, 15:55:37  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 25, 2004, 15:53:27 ; Search time 21 Seconds

(without alignments)  
2216.986 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644

Sequence: 1 MPESLALPNDPEWGFATPAAY.....KSALVLKDMFARQVKVA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR78: \*  
2: PIR1: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1037	39.2	510	2 F86392	TIK7.7 protein - A
2	1003	37.9	577	2 T47838	beta-glucosidase-1
3	999	37.8	577	2 T02401	probable beta-gluc
4	984.5	37.2	591	2 T02402	beta-glucosidase h
5	983	37.2	551	2 T09647	beta-glucosidase (
6	976	36.9	509	2 A57512	beta-glucosidase B
7	965	36.5	560	2 T02404	probable beta-gluc
8	964.5	36.5	562	2 S78039	flavonol glycosyl
9	960.5	36.3	534	2 T47836	beta-glucosidase-1
10	944.5	35.7	517	2 T02403	probable beta-gluc
11	943.5	35.7	549	2 T09657	beta-glucosidase (
12	932	35.2	574	2 S50756	beta-D-glucosidase
13	928	35.1	547	2 JCT539	beta-glucosidase (
14	927	35.0	519	2 S45723	beta-glucosidase (
15	925.5	35.0	506	2 T02400	probable beta-gluc
16	922	34.6	527	2 T02128	beta-glucosidase h
17	914	34.6	514	2 T47837	beta-glucosidase-1
18	913.5	34.5	565	2 T14732	probable beta-gluc
19	902	34.1	563	2 T02770	beta-D-glucosidase
20	896	33.9	531	2 S23940	beta-glucosidase (
21	884	33.4	566	2 A48860	beta-glucosidase (
22	880.5	33.3	493	1 GJYJ31	beta-glucosidase (
23	873	33.2	524	2 H96687	probable beta-gluc
24	873	33.0	1920	2 S43721	lactase (EC 3.2.1.
25	868	32.8	524	2 G96687	probable beta-gluc
26	863.5	32.7	524	2 S57621	thioglycosidase (E
27	863.5	32.7	1918	2 S43719	lactase (EC 3.2.1.
28	859	32.5	520	2 T02127	beta-glucosidase h
29	858.5	32.5	548	2 S26149	thioglycosidase (E

30	857.5	32.4	1927	2 S01168	beta-glycosidase c
31	854	32.3	536	2 S56554	thioglycosidase (E
32	852.5	32.2	1926	2 S01169	beta-glycosidase c
33	851.5	32.2	1928	2 JS0610	beta-galactosidase
34	849.5	32.1	541	2 S56653	thioglycosidase (E
35	845	32.0	544	2 S19149	thioglycosidase (E
36	843	31.9	531	2 T51956	probable beta-gluc
37	839	31.7	446	2 S34570	beta-glucosidase (
38	831.5	31.4	544	2 S39550	thioglycosidase (E
39	831.5	31.4	547	2 S56656	thioglycosidase (E
40	826	31.2	528	2 C96564	probable beta-gluc
41	825	31.2	575	2 S43128	beta-D-glucosidase
42	825	31.2	882	2 B96788	protein T4012.15 l
43	819	31.0	535	2 T52048	probable beta-gluc
44	814	30.8	496	2 G96516	hypothetical prote
45	813.5	30.8	447	2 C83890	beta-glucosidase b

## ALIGNMENTS

## RESULT 1

TIK7.7 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 27-Nov-2001

C:Accession: F86392

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hultzer, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F86392

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-510 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:9797746; PIDN:AAF98564.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: Agrobacterium beta-glucosidase

Query Match 39.2%; Score 1037; DB 2; Length 510;

Best Local Similarity 44.1%; Pred. No. 4.3e-74;

Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

QY	6	ALPNDPEWGFATPAAYIQEGAVKGGKGRPSIWDYCHLEPRTGANGDVACHYHRDED	65
DB	36	SPFGKGFVFGTASAPFHQEGAVKAEGRGPTIMDFSHFGKITPESNADVVDVDDHRRFED	95
QY	66	FDLLTKYGAAYRPSLSMSRIIPLGKRLDPVNEGIEFYSKLDALLRSGITPMVTLTYH	125
DB	96	VQLMKMGMDAYRFSISWIRIFPNG--VGHINAGIDHVKLNALLAKIIEPVTLTYH	153
QY	126	DLPOALHDIRGYGLWLVNEVQLDERYARLCFERFGGRVQVMITINXPMWIAIYATGSN	185
DB	154	DLQALHDIRGLWLVNPIIN--DFAVAVEVCFQFGGRVQVMITINXPMWIAIYATGSN	212
QY	186	APGRSSINKIST--EGNTATEPWLAKQAQMSHARAVAVSRPFRSQKQIGISLNGDY	243
DB	213	APGRCTILFRLTGREGNSSTPEYVGHNVLTATVSDIRKRYKAKQGGSLGIAPDPVM	272
QY	244	YEPWDSNEPRDKAERBMEFHTGWFANPITLKKDYPESKXQGLERLPAITPADPAIIN	303
DB	273	FEP-ESKTEDIDRAQDAQDFQGLWFLDPLMF--GDYPSRSKRSVGRSLPFTSGSQSLVK	330
QY	304	AGETDFGMNYYTSOPARHLIDGVPETDYLGAIEHQENKDG-----SPVGEBSGL	354

Db 331 -GSLDVGINHTTYAARN----NATNLIGTLHDVADSGVTLLPFKGLSTIGDRASS 384  
QY 355 AMLRSCPMDFRKLARVYGLYK-PYITENGCPGCEENMTCEAVNDPFRIRYFDSHL 413  
Db 385 IMLYIPRGMRSLMNYIKRYGNPVPFITENGMDDPNLSLRKALNDAKKIKKHHDYL 444  
QY 414 DSISRLITDDGVVVKGYFAMALLDNLEWSDGYGPRGVTFTDY-TTLKTPPKS 466  
Db 445 SSLQASIKEDGCVNKGYPWMSLLDNWEMAGYSRFGLYFVDYRDLKRYPKDS 498

## RESULT 2

T47838  
beta-glucosidase-like protein - Arabidopsis thaliana  
N/Alternate names: protein T209.120  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 21-Jul-2000  
A/Accession: T47838  
R/Nyatura, G.; Partam, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A/Reference number: Z24475  
A/Accession: T47838  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-577 <NVA>  
A/Cross-references: EMBL:AL138658  
A/Experimental source: cultivar Columbia, BAC clone T209  
C/Genetics:  
A/Map position: 3  
A/Intons: 45/3; 69/1; 87/3; 113/1; 139/1; 168/2; 293/2; 366/1; 377/3; 412/1; 448/2  
A/Note: T209.120  
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 37.9%; Score 1003; DB 2; Length 577;  
Best Local Similarity 43.7%; Pred. No. 2.5e-71;  
Matches 205; Conservative 77; Mismatches 175; Indels 12; Gaps 8;  
QY 6 ALPNDFEMGPAATAVQIEGAVKEGGRGSIWDYCHLEPSTNGANGVACDHYRDEDF 65  
Db 30 SPDDFIFIGTASAPQYEGATSEGGKSPITWDHFSFLYPTERTKMNADVAFYHRYKDD 89  
QY 66 FDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHM 125  
Db 90 IKLMEIMMDAPRFSISMSRLIPSGKLDGVNKEGVQYKDLIDELLANDIQPSMTLYHM 149  
QY 126 DLPOALHRYGMLNVEEVDLPFRYARLCERFGRDVQNNITINXPMWIOAIYATGSN 185  
Db 150 DHPQSLDEYGGFLSPKIVE-DFRDFARICFEFEFGDKVMWTTINEPYIMTVAGYDGNK 208  
QY 186 APGRSS--INHGSTEGRNTATEPMLAGKAQIMSHARAVAVSRDPRPSQKQIGISLNGDY 243  
Db 209 AAGRCSSKVNNEKCGAGDSSTEPYIVSHHTTLAAHAAVEEFKCKTSHDQIGIVLSPRW 268  
QY 244 YEPWDSNEPRDKEAERMEFHIGMFANPIFLKDYPESMKQGLERLPALTTPADFAIIN 303  
Db 269 FEPYHSDTDDEKAERLAFELIGMHLDPV-IHGDIYELIVKVGKGLPSTTVEQSCKLQ 327  
QY 304 AGEDFYGMNTYQOPARHLDPVPETDYLGAIHEHQB---NKQSPVGEESGLAWLRS 359  
Db 328 -NSSDFGGINYYTARFAAHPHIDPEKPRFCTDH-HYEMKLTJNSGHIIGCGEERGLFS 385  
QY 360 CPDMFRKLARVYGLYK-PYITENGCPGCEENMTCEAVNDPFRIRYFDSHLDSISK 418  
Db 386 HPEGLRKLAVNIKRYNNMYYIKENGINDNDGTRKEELVKDTPFERIEYHKTFFELHK 445  
QY 419 AITDDGVVVKGYFAMALLDNLEWSDGYGPRGVTFTDY-TTLKTPPKS 466  
Db 446 AIVEDGCDVGVYAMSLDNFEMEGHTARGLYVYDFVNGLKRYPKDS 494

RESULT 3  
T02401  
probable beta-glucosidase At2g44460 [imported] - Arabidopsis thaliana

N/Alternate names: beta-glucosidase homolog F411.27  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
A/Accession: T02401; G84878  
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,  
submitted to the EMBL Data Library, May 1998  
A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.  
A/Reference number: Z14667  
A/Accession: T02401

A/Status: translated from GB/EMBL/DBD  
A/Molecule type: DNA  
A/Residues: 1-577 <ROU>  
A/Cross-references: EMBL:AC004521; NID:g3128166; PID:g3128168  
A/Experimental source: cultivar Columbia  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Varakken, S.E.; Umayam, L.; Tallon, L.;  
Euse, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: G84878  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-577 <STO>  
A/Cross-references: GB:AE002093; NID:g3128188; PIDN:AA016092.1; GSPDB:GN00139  
C/Genetics:  
A/Map position: 2  
A/Intons: 48/3; 72/1; 90/3; 116/1; 142/1; 171/2; 256/3; 295/2; 376/1; 409/1; 445/2  
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 37.8%; Score 999; DB 2; Length 577;  
Best Local Similarity 42.9%; Pred. No. 5.2e-71;  
Matches 204; Conservative 78; Mismatches 184; Indels 10; Gaps 8;  
QY 8 PNDPFGFRTAAVQIEGAVKEGGRGSIWDYCHLEPSTNGANGVACDHYRDEDF 67  
Db 35 PDNFVFGTASAPQYEGATSEGGKSPIDYSHPTERTKMNADVAFYHRYKDDIK 94  
QY 68 ILTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYHM 127  
Db 95 IKLMEIMMDAPRFSISMSRLIPSGKLDGVNKEGVQYKDLIDELLANDIQPSMTLYHM 154  
QY 128 POALHRYGMLNVEEVDLPFRYARLCERFGRDVQNNITINXPMWIOAIYATGSNAP 187  
Db 155 PQSLDEYGGFLSPKIVE-DFRDFSRVCFEFGDKVMWTTINEPYIMTVAGYDGNKAV 213  
QY 188 GRSS--INHGSTEGRNTATEPMLAGKAQIMSHARAVAVSRDPRPSQKQIGISLNGDY 245  
Db 214 GRCSKVNNSKCGGDSSTEPYIVSHHTTLAAHAAVQEF-RKCNKTQDQIGIVLSPLWFE 272  
QY 246 PWDNSNEPRDKEAERMEFHIGMFANPIFLKDYPESMKQGLERLPALTTPADFAIIN 305  
Db 273 PYVSASPPANNEAVKRALATELDMHLDPV-IHGDIYEMKGLAGNRLPSTTPQSKLK-N 330  
QY 306 ETDYFGMNTYQOPARHLDPVPETDYLGAIHEHQBNDKQSPVGEESGLAWLRS 365  
Db 331 SSDFGGINYYTARFAAHPIQADPARPFTVDHQLQWRGKIYANVINIRGI--LQSHPEGLR 388  
QY 366 KHLARVYGLYKPI-YITENGCPGCEENMTCEAVNDPFRIRYFDSHLDSISK 418  
Db 389 KALNIYKDKTNNPIYIKENGINDYDDGTSKEELNTPFRISYHEDHQOQKAIIDBG 448  
QY 425 VVVKGYFAMALLDNLEWSDGYGPRGVTFTDY-TTLKTPPKSALVLDKMFAROR 479  
Db 449 CVVRGYVWMSLLDNFEMEGHTARGLYVYDFVNDNLTLIPKDSVWVFQFLDYKAK 504

RESULT 4  
T02402  
beta-glucosidase homolog At2g44470 [imported] - Arabidopsis thaliana  
N/Alternate names: beta-glucosidase homolog F411.28  
C/Species: Arabidopsis thaliana (mouse-ear cress)



Query Match 36.9%; Score 976; DB 2; Length 509;  
Best Local Similarity 43.0%; Pred. No. 2,9e-69;  
Matches 206; Conservative 81; Mismatches 170; Indels 22; Gaps 12;

QY 8 PNDFFMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGDVACDHRYDEDD 67  
D 45 PAFGVGTASAAAYOVEGMAKGGRCICWDVAALQMGATGADYVDEHRRKEDVG 104  
QY 68 LUTYKAKAYRPSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRRGITTPWTLVHDL 127  
D 105 IMKMGFEFDYRPSISMSRIIPDGG--TGKYNQGVYNNLIDYMLQQGITPYANNLYHDL 162  
QY 128 PQLADHYRGMLNVEEVDQDFEERARLCEFRGDRVQNMWITIXXPMQAIYATAGSNAP 187  
D 163 PLALHQOYLQWLSPKIVGA--FADYAEFCFKVGDYKAKWFTNERVVAALGTDGHPAP 221  
QY 188 GSSSINKSTEGNTATEPMLAGKAOIMSHARAAYVSRDFRPSQKQIGISLNGDYEPW 247  
D 222 GRCS--KCPAGGDSRTEPYIVTHNIIILSHAAVQGRREKYOPHQGRIGILLDFVWYEP 278  
QY 248 DSENEPRDEKAEARRMEFHIGWFANPIFLKQYBESMKQLGERLPLTPADPAIINAGET 307  
D 279 HSDTPADDAQAARADPFHIGWFLDPL--TNGRYPSMLKIVGNRLPGFS--ADESRWVKSI 336  
QY 308 DFYGMNYYTTSOPARHLDG---PVPETD--YLGAIHHEQENKDGSPVGEESGLAWLRSCP 361  
D 337 DYVGINQYTSYMKDPGAMNQTPVSGYQDDWYGFYE---RNGVPIOPRANSMDLYIVP 392  
QY 362 DMFRGHRLARYGLYKRP--IYITENCCPCGGEENMTCEAVNDPFRIRYFSDHLSISKAI 420  
D 393 WGMNAAYVYKERYNPNMILISENMDPG--NVSIAQGVHDYAIRYRYDYITELKKA 450  
QY 421 TQDGVVVGYPAMALLDNLLEMSDGYGPRFGVTFDYTTLKRPKPSAIVLDMWFAARQ 479  
D 451 DNGARVAGYFAMSLDNLFEWRLGYTARFGIYVYDNTLKYPKDSALMFKNMSEKR 508

## RESULT 7

T02404  
probable beta-glucosidase homolog F411.30 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
C/Accession: T02404; B84879  
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, May 1998  
A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.  
A/Reference number: Z14667  
A/Accession: T02404  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-560 <ROU>  
A/Cross-references: EMBL:AC004521; NID:g312816; PIDN:AAIC16095.1; PID:g3128191  
A/Experimental source: cultivar Columbia  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402:761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: B84879  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-560 <STO>  
A/Cross-references: GB:AE002093; NID:g3128191; PIDN:AAIC16095.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2g44490; F411.30  
A/Map position: 2  
A/Intons: 3/2; 3/3; 76/3; 102/1; 157/2; 242/3; 281/2; 354/1; 400/1; 438/2  
C/Superfamily: Agrobacterium beta-glucosidase

Query Match 36.5%; Score 965; DB 2; Length 560;  
Best Local Similarity 42.3%; Pred. No. 2,4e-68;

Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDFFMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSR--TNGANGDVACDHRYDE 64  
D 18 SFRKGLFEGTASSSYQYEGAVNCAKQSVWDHFSNFPRIIDSSGNAVVPYHKKYK 77  
QY 65 DFDLLTYKAKAYRPSLSMSRIIPLGRLDPVNEEGIEFYSKLIDALLRRGITTPWTLVH 124  
D 78 DIKMKDINNDSFRLSIAMPRLVYKGRDRGVEEGIKFYNDVIDELLANEIRPLVTIFH 137  
QY 125 WDLFQALHDYRGMLNVEEVDQDFEERARLCEFRGDRVQNMWITIXXPMQAIYATAGTS 184  
D 138 WDIQDLEDEYGFSL--EQLIIDFDYASLCEFRGDRVSLWCTMNPWYVAAGYDTGA 196  
QY 185 NAFGRSS--INKSTEGNTATEPMLAGKAOIMSHARAAYVSRDFRPSQKQIGISLNGD 242  
D 197 KAPRCISKYVNGASVAGMSGYEAIVYSHMMLAHAEVVF--RKCHINKQIGIAHNP 255  
QY 243 YFERWDSNEPRDEKAEARRMEFHIGWFANPIFLKQYBESMKQLGERLPLTPADPAI 302  
D 256 WYEPYDPSDDVDEGCNRAMDFMLGMHOFYAC--GDYPTMKSVGDRLPSTPEQSKL 314  
QY 303 NAGETDFYGMNYYTTSQFA--RHLDGPVP--ETDYLGAIHHEQENKDGSPVGEESGLAW 357  
D 315 -ISGCVYGINYSSLFVYSIKHVDPTQPTKRTDQ--GYDMKNTIDGKQIAKQGSSEWS 371  
QY 358 RSCPDMFRKRLARYGLYKRP--PIYITENCCPCGGE---ENMTCEAVNDPFRIRYPDS 411  
D 372 FYTPGTJRLNLKVKKTYGNPILITENGY---GEVAEQSLYMNPSIDTERLEYIEG 428  
QY 412 HLDSIKAITQDDVVVVGYPAMALLDNLLEMSDGYGPRFGVTFDYTTLKRPKPSAIVL 470  
D 429 HIAHQALHEDGVREGYVWSLNDLFEMNSGYGRVGLYYIDYKDGILRRYKMSALWL 488  
QY 471 KD 472  
D 489 KE 490

## RESULT 8

S78099  
furostanol glycoside 26-O-beta-glucosidase F26G - Costus speciosus  
C/Species: Costus speciosus  
C/Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 20-Jun-2000  
C/Accession: S78099; S68626  
R/Inoue, K.; Shibuya, M.; Yamamoto, K.; Ebizuka, Y.  
submitted to the EMBL Data Library, January 1996  
A/Description: Molecular cloning and bacterial expression of a cDNA encoding furostanol  
A/Reference number: S78099  
A/Accession: S78099  
A/Molecule type: mRNA  
A/Residues: 1-562 <INO>  
A/Cross-references: EMBL:D83177; NID:g1374990; PIDN:BA11831.1; PID:g1374991  
A/Experimental source: CSF26G1; rhizome  
R/Inoue, K.; Ebizuka, Y.  
FEBS Lett. 378, 157-160, 1996  
A/Title: Purification and characterization of furostanol glycoside 26-O-beta-glucosidase  
A/Reference number: S68626; MUID:96140730; PMID:8549824  
A/Accession: S68626  
A/Molecule type: protein  
A/Residues: 84-86, 'AE', '89-100', 'N', '102-103 <INW>  
A/Experimental source: rhizome  
A/Complex: heterodimer; 54k chain and 58k chain  
C/Function:  
A/Description: catalyses the removal of C-26-bound glucose moiety of furostanol glycoside  
A/Note: strongly inhibited by diosgenin  
C/Superfamily: Agrobacterium beta-glucosidase  
C/Keywords: heterodimer

Query Match 36.5%; Score 964.5; DB 2; Length 562;  
Best Local Similarity 43.5%; Pred. No. 2,7e-68;  
Matches 204; Conservative 78; Mismatches 168; Indels 19; Gaps 12;



QY 6 ALPNDPEWGFATTAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64  
 Db 95 SFPGRGTFGAASAAYOIEGAVKEGGRPSIMDTFTTHDPEKIDHSGDKATOSYKYYKE 154  
 QY 65 DFDLLTYGAKAYRFSISMSRIIPLGRLDPVNEBEGIEFYSKIDALLRGITPWTLYH 124  
 Db 155 DVLLKXLDGDSYRFSISMSRIIPKGTLOGGNGEGIQYVNDLINELKNGIRPWTLYH 214  
 QY 125 WDLPOLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 184  
 Db 215 WDVPQALDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 273  
 QY 185 NAVGRSSINHGSTEGRATEPMLAGKQIMSHAAVAVYSGDRFSPKOGQIGISLNDY 244  
 Db 274 HAAGRGSTWYGCAPGADANPEYETHNLLAHANAVALYDNYKATONGELITLNSLWY 333  
 QY 245 EPPDSDNEPRDKEAERMEFHIGFANPILKODPESMKKQJGERLPALTPADFAILMA 304  
 Db 334 EPY-SKSHEDVEATRALDPMFGWMDPL-VNGDYPFIMRALYDRLPFFTHAESELIK- 390  
 QY 305 GETDFYGMNYYTSGFAHLDGPV-----PENTDYGAIHEROENKDSPVGEESGLAMLR 359  
 Db 391 GSYDFIGINYYTSSYAOH--APVTEHDPDNSYFDYSVNGSGEKNGVIGPLOG-SWITYF 447  
 QY 360 CPDMFRKHLARVYGLYKGP-IYITENGCPGCEENMTCEBAVNDPFRIRYFDSHLSISK 418  
 Db 448 YPFGGLKLLLVKRYKRNPKIYITENG---TAIEKKEKGVPLHDPKEKELITHTLAQVLQ 504  
 QY 419 AITODGVVVGYPAMALLDNLWSDYGPFGVTFTDY-TTLKRTPKKS 466  
 Db 505 AI-REGVAVGHFWALTDFEMDKYTERFGIYIDYDKNFQPMKDS 552

## RESULT 9

T47836  
 beta-glucosidase-like protein - Arabidopsis thaliana  
 N/Alternate names: protein T209.100  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
 C/Accession: T47836  
 R/Vyakarna, G.; Partmann, B.; Dauner, D.; Steer, W.; Holland, R.; Weishejgartner, M.;  
 submitted to the Protein Sequence Database, February 2000  
 A/Reference number: 224475  
 A/Accession: T47836  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-534 <NTA>  
 A/Cross-references: EMBL:AL138658  
 A/Experimental source: cultivar Columbia; BAC clone T209  
 C/Genetics:  
 A/Map position: 3  
 A/Intons: 51/1; 95/1; 121/1; 150/2; 235/3; 274/2; 347/1; 393/1; 429/2  
 A/Note: T209.100  
 C/Superfamily: Agrobacterium beta-glucosidase

Query Match 36.3%; Score 960.5; DB 2; Length 534;  
 Best Local Similarity 42.8%; Pred. No. 5.2e-66;  
 Matches 207; Conservative 75; Mismatches 185; Indels 17; Gaps 11;  
 QY 8 PNDPEWGFATTAAYOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRDE 67  
 Db 14 PEGFLFTASSAYOIEGABARNDARPGESVMDTFVVKYERNYCYNADQAIERYNNYKODIQ 73  
 QY 68 LRTYGAAYRFSISMSRIIPLGRLDPVNEBEGIEFYSKIDALLRGITPWTLYH 127  
 Db 74 RMDINNDARFISMSRIIPLGRLDPVNEBEGIEFYSKIDALLRGITPWTLYH 133  
 QY 128 POLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 187  
 Db 134 POLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 192  
 QY 188 GRSS--INKHSTEGNTATEPMLAGKQIMSHAAVAVYSGDRFSPKOGQIGISLNDY 245

Db 193 GRASKYMNBAVAVGESGLVYTVSHNLLAHAEAVEF-RNNPKCKDKIGIAHCPWFE 251  
 QY 246 PMSDNEPRDKEAERMEFHIGFANPILKODPESMKKQJGERLPALTPADFAILMA 305  
 Db 252 PYDSNCKDIEACERAMERFEGHMDPT-VYGYPAVMKKSICKRLPSFPAOSKLR-G 309  
 QY 306 ETPFYGMNYYTSGFAHLDGPV-----DGPVETDYGAIHEROENKDSPVGEESGLAMLR 360  
 Db 310 SFPFGVGNYSAYVYKNIDVNHDKPNMSD--ARLEMKRENNAGQTLGVRGSEMDFLY 367  
 QY 361 PDMFRKHLARVYGLYKGP-IYITENGCPGCEENMTCEBAVNDPFRIRYFDSHLSISK 418  
 Db 368 PDMFRKHLARVYGLYKGP-IYITENGCPGCEENMTCEBAVNDPFRIRYFDSHLSISK 426  
 QY 419 AITODGVVVGYPAMALLDNLWSDYGPFGVTFTDY-TTLKRTPKKS 466  
 Db 427 AITODGVVVGYPAMALLDNLWSDYGPFGVTFTDY-TTLKRTPKKS 474  
 QY 478 QRYK 481  
 Db 487 EEIE 490

## RESULT 10

T02403  
 probable beta-glucosidase At2g44480 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 02-Feb-2001  
 C/Accession: T02403; A84879  
 R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,  
 submitted to the EMBL Data Library, May 1998  
 A/Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.  
 A/Reference number: Z14667  
 A/Accession: T02403  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-517 <ROU>  
 A/Cross-references: EMBL:AC004521; NID:G3128166; PID:AAIC16094.1; PID:G3128150  
 A/Experimental source: cultivar Columbia  
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;  
 euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A/Reference number: A84420; MIMD:20083487; PMID:10617197  
 A/Accession: A84879  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-517 <STO>  
 A/Cross-references: GB:AE002093; NID:G3128190; PID:AAIC16094.1; GSPDB:GN00139  
 C/Genetics:  
 A/Gene: At2g44480; F411.29  
 A/Map position: 2  
 A/Intons: 54/3; 78/1; 97/3; 123/1; 149/1; 178/2; 262/3; 301/2; 374/1; 384/3; 419/1; 455/  
 C/Superfamily: Agrobacterium beta-glucosidase

Query Match 35.7%; Score 944.5; DB 2; Length 517;  
 Best Local Similarity 41.1%; Pred. No. 9.1e-67;  
 Matches 197; Conservative 91; Mismatches 160; Indels 31; Gaps 11;  
 QY 6 ALPNDPEWGFATTAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64  
 Db 39 SFPDPRFGAASSAYOIEGAVKEGGRPSIMDTFTQYPERKIDSGNSGVADSEFYRFXE 98  
 QY 65 DFDLLTYGAKAYRFSISMSRIIPLGRLDPVNEBEGIEFYSKIDALLRGITPWTLYH 124  
 Db 99 DVAMKEIGDSEFRFISMSRIIPRGTVAGVQAQAGINFNHILINELISNGIRPLVTLF 158  
 QY 125 WDLPOLHDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 184  
 Db 159 WDVPQALDEYGGMLNVEEYQDPERYARLCFERFGDRVONMTINXPMWQAIYGVATGS 217  
 QY 185 NAPGR--SINKHSTEGNTATEPMLAGKQIMSHAAVAVYSGDRFSPKOGQIGISLNDY 243

```
Db      218 IAPGCSSTVQNCYGTGNSATPEPYLVAHVLILSHATVQLYRRKQSFHGCTIGMTIQYTW 277
      244 YEPMSNBRDKEAERMEFHIGWFPANPILFKKQYPSKMKQLGERLPALTPADPAIIN 303
      278 MTP-KYNTFPACREAAKRRALDFFPGWFADPI-TVGYGPKTMRBLVGNRLPKFTKQSKXVR 335
      304 AGETDYGKNNYTSQFARHL-----DGVPEPDYDGLAHEHOENKDGSPVGE 351
      336 -GSFPFGIANTTSKYVEDVFWYATNTNLSYTTDSKVNQT-----TEKNGVPEBP 384
      352 SGLAMLRSCPDMPFRKLARVGLYKRP-IYTTNGCPCPGEBNMTCEAVNDPFRIRYFD 410
      385 TSADWLFTCPCEGFQOVLLIYKSKFQNPVILVTENGMPESENKSLSVNIALNDEAKIKYHQ 444
      411 SHLDSISKATDGGVGVVGFAMALLDNLNEMSDGPRPGVTFPTYTT-LKRTPKSKAL 468
      445 LHLTLLLEAVSO-GADVKGYYIWSLMDPEWMEFGYKRYGLVYDFQDGLKRHLKSSAL 502
```

## RESULT 11

```
T09657
beta-glucosidase (EC 3.2.1.21) precursor, cyanogenic - black cherry
N:Alternate names: amygdalase; cellobiase; gentiobiase
C:Species: Prunus serotina (black cherry)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
R:Poulton, J.E.; Jurrk, S.
  submitted to the EMBL Data Library, February 1996
A:Reference number: Z16802
A:Accession: T09657
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-549 <POU>
A:Cross-references: EMBL:U50201; NID:G1236960; PID:G1236961
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycosidase; hydrolase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-549/Product: beta-glucosidase, cyanogenic #status predicted <MAT>
```

```
Query Match      35.7%; Score 943.5; DB 2; Length 549;
Best Local Similarity 42.5%; Pred. No. 1.2e-66;
Matches 198; Conservative 84; Mismatches 169; Indels 15; Gaps 11;

      11 FEMGPAATAVOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYRYDEFDLL 69
      54 FTFGAATAAYOLEGANIDGRPSVDMFTHEHPEKIDSGNGVAILDQYHRYKEDVALIM 113
      70 TKYGAKAARFSLMSRIITPLGRLDPVNEBGEIEFYSKLIDALLRGITTPWVTLYHMDLPQ 129
      114 KDMGLDAVRFISWSRLLPNTLSGGINKGIEYYNNLTNELLNGLIEPLVTLFPHMDVPQ 173
      130 ALHRYGGMVLNVEEYOLDPERFYARLCFERPGDRVONMTTINXPMIOAIYGAATGSNA 189
      174 ALVVEYDGLSLPRIVD-DFEAYANLCKYEPGDRKAKMTTLNEPTVSNHGTTIGIHA 232
      190 SSINKHST--EGNTATEPWLAKGAQIMSHARAVAVYSHDRPSSQKGQIGISLNGDY 247
      233 CSCWDPYTCGLGDSGTBEYLVTHHLLAHLAAVKLREKYQASONGVIGITIVSHWEP 291
      248 DSNBPRDKEAERMEFHIGWFPANPILFKKQYPSKMKQLGERLPALTPADPAIIN 307
      292 ASBEOQDDKDAASRALDFFWGFMEPL-TRGDYPTMISIVGSRLLPNFTBEGSKSLN- 349
      308 DFGYGNYYTSQFA----RHLDGPVPEYDYLGA-IHEHOENKDGSPVGEESGLAMLRSCPD 362
      350 DYICGVYTSARVAYATNNVSVFPPPSVATDAYVNTTDLNGVPIBPQASMLYLYPK 409
      363 MFRGHARVGLYKRP-IYTTNGCPCPGEBNMTCEAVNDPFRIRYFDLHDSISKAIT 421
      410 GLYDLVLYTEKXNDPVMYITENGMDEFNPKLSLEBALDANRIIDYYRHLCYLQA 468
      422 QDGVVVGVGFAMALLDNLNEMSDGYPGRGVTFTTYTT-LKRTPKKS 466
```

```
Db      469 KEGANVOGYFAMSLLDNLFMSSEGYTTFVFGIYNDYNGLERHSKLS 514
```

## RESULT 12

```
S50756
beta-D-glucosidase - oat
C:Species: Avena sativa (oat)
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
R:Gus-Mayer, S.; Brunner, H.; Schneider-Poetsch, H.A.W.; Ruediger, W.
  plant Mol. Biol. 26, 909-921, 1994
A>Title: Avenacosidase from oat: purification, sequence analysis and biochemical character
A:Accession: S50756; MUID:95093031; PMID:8000004
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-574 <GUS>
A:Cross-references: EMBL:X78433; NID:G505278; PID:CAAS5196.1; PID:G505279
C:Superfamily: Agrobacterium beta-glucosidase
```

```
Query Match      35.2%; Score 932; DB 2; Length 574;
Best Local Similarity 40.9%; Pred. No. 1e-65;
Matches 202; Conservative 86; Mismatches 182; Indels 22; Gaps 14;
```

```
      1 MPESIALPNDFEMGPATAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHY 59
      68 VPKRDMFPPEFMGAASAAVOIEGANMEGKSGSSMDFCHSPDDIMDSNDVAAASTY 127
      60 HRYDEFDLLTKYGAKAARFSLMSRIITPLGRLDPVNEBGEIEFYSKLIDALLRGITTPW 119
      128 YMYKEDVRMLKEIGMDSYRFSISMPRLPKGTLDGGINHEGIQYNDLDCLENGIKFY 187
      120 VTLXHMDDLPAALDRYGMVLNVEEYOLDPERFYARLCFERPGDRVONMTTINXPMIOAIY 179
      188 ITLFHWDTPOALDEXKDFLD-RIYKDYDYDVATVCFEHRGDKYKWFTEHPSFCGLG 246
      180 YATGSNAPG-RSSINKHST--EGNTATEPWLAKGAQIMSHARAVAVYSHDRPSSQKGQIG 236
      247 YGTGHAHAPGARGCAGMTCVIPEEDALRNPIYVHNLHLAAEVYDVYNNK-FYGGDGOIG 305
      237 ISLNGDYEPWDSNEPRDKEAERMEFHIGWFPANPILFKKQYPSKMKQLGERLPALTP 296
      306 MVLDVVAVEPYGNPF-LDQQAQERAILDFHIGFLBPM-VRGDYPFSMRSLVGRPLPFTK 363
      297 ADAFALLNAGRTDYGKNNYTSQFARHLD-----GVPPEPDYDGLAHEHOENKDG-- 349
      364 SEQEKLVSSYDVFVGINYYTSRPAKIDISPEFIPKINTD--DVSINBPEVNDNGIPIG 419
      350 EESGLAMLRSCPDMPFRKLARVGLYKRP-IYTTNGCPCPGEBNMTCEAVNDPFRIR 407
      420 PDVGWYFIYSYRGLKNILLRMKEKYGNPPIYITENGTADMOMGMPRMTDPLDDLRLE 479
      408 YFPHSLDLSKAITDQGVVYKGFAMALLDNLNEMSDGYPGRGVTFTTYTT-LKRTPKKS 466
      480 YLQOHTALKEADLRRLRGHFTSLIDNFEWSIGYRFGIYVIDRNDGCKRIMKKS 539
      467 ALVLKDMFAPAROV 480
      540 AKMLKEPFGATKYL 553
```

## RESULT 13

```
JC7539
beta-glucosidase (EC 3.2.1.21) - Thai rosewood
N:Alternate names: dalcocchinase; dalcocchinin-8'-O-beta-glucoside beta-glucosidase
C:Species: Dalbergia cochinchinensis
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
R:Calins, J.R.K.; Champatnatchai, V.; Sriomsap, C.; Wattman-Liebold, B.; Thiede, B.; St
  J. Biochem. 128, 999-1008, 2000
A>Title: Sequence and expression of Thai rosewood beta-glucosidase/beta-fucosidase, a far
A:Reference number: JC7539; MUID: 21062552; PMID:11098143
```

A:Contents: Seeds  
A:Accession: J07539  
A:Molecule type: mRNA  
A:Residues: 1-547 <CAI>  
A:Cross-references: GB:AF163097  
A:Accession: F07111  
A:Molecule type: protein  
A:Residues: 66-74;81-95;100-104;106-112;127-141;138-351;410-422;465-472;500-506;507-516  
A:Comment: This enzyme belongs to family 1 glycosyl hydrolase glycoproteins and plays a role in degradation, and defense against herbivores. It has aglycone specificity for isoflavonone  
C:Superfamily: Agrobacterium beta-glucosidase  
C:Keywords: glycoprotein; glycosidase; growth regulation; hydrolase

Query Match	35.1%	Score 928;	DB 2;	Length 547;
Best Local Similarity	42.1%	Pred. NO. 2e-65;		
Matches 204;	Conservative	78;	Mismatches 173;	Indels 30;
				Gaps 12;

Qy	8	PNDPEMGATAYQJEGAKVEGRGSI	INDYTCHELPSS-TNGANDVACDHHRJDEDF	66
Db	46	PSDFIFGJASSSYQJEGE----	GRVSIYDNTTTHQVPEKIKADRSNDVAVDQHRKKDI	101
Qy	67	DLITRYKAKAYRFSLSWSRIIPLGGR	LDPVNEBGIEFFYSKLIDALLRGITTPVTL	126
Db	102	AIMQDMNLDAAYMSISWPRILPTGR	VSGGIGNQTDVYNRLLINESLANGITTPVTL	161
Qy	127	LPQALHDERYGGWLJANEVVO	LDPERYARLCFERRGDRVQMMITINXPMIAIYVATGNSA	186
Db	162	LPQALHDERYGGFLN-HSVYND	RQDADLCFQLEFGRVKKMITINESIIFRANGYAGMFA	220
Qy	187	PGR--SSJNKSTEGCNTATEPW	LACKAQIMSHARAVAVYSRDRP	PSQKQIGISLGDYY 244
Db	221	PGRCSPSYNPCTCGDAGET	YLVAHNLLISHAATQVYKRYKQEHQKGTIGISLHVWV	280
Qy	245	EPWMSNEPRDKEAARRMF	FHJGWFPANPFIKDDVPESSKKQGERL	PALTPADFAILNA 304
Db	281	IPL-SNSTDQNAOTRYLIDFT	CGWENPDLTAGR-YPDSQYLVGDMLPKFT	-TDQAKLVK 337
Qy	305	GETDPYGMNYYTSQFARHL	DGPV-----PETDYICAIHEHQBKDGSPVGEESGLA	355
Db	338	GSPFICIGINYTTNATYKSDA	TCCPPSYLTQPOVTL-----QQRNGVFI	GPVTPSG 390
Qy	356	WLRSQPMDFRKLAVVGLY	GKRP-LYITENGCPRCEEMWTC	CEAVNDPRIRIYFPSHLD 414
Db	391	WMCITPKLRDILLFYKEK	RKNPNPLVITTENGIDKNDASLSEBSL	LDYIRIDISYRHLF 450
Qy	415	SISKAITODGVVVKQYFAM	ALLDNDLEWSDGYPFRGVTFTDYTL	KRTPKSAALVYKDMF 474
Db	451	YVRVYAI-RSGANVAKGFAM	SLDNDFEMAGYTSRFGLYFVN	YTTLMRXPRLSATWPK-YF 508
Qy	475	AAROR 479		
Db	509	LARQ 513		

```

RESULT 14
S45723
P60 protein - oat
C:Species: Avena sativa (oat)
C:Date: 10-Dec-1994 #sequence_revision 19-Apr-1996 #text_change 17-Mar-1999
C:Accession: S45723
R:Gus-Meyer, S.; Brunner, H.; Schneider-Poetsch, H.A.W.; Lotsepelch, F.; Ekerskorn, C.;
FEBS Lett. 347, 51-54, 1994
A:Title: The amino acid sequence previously attributed to a protein kinase or a TCPI-rel
A:Reference number: S45723; MUID:94283628; PMID:8013661
A:Accession: S45723
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-519 <GUS>
;Superfamily: Agrobacterium beta-glucosidase

```

Query Match	35.1%	Score 927	DB 2:	Length 519;
Best Local Similarity	40.7%	Pred. No. 2.2e-65;		
Matches 201; Conservative	89;	Mismatches 182;	Indels 22;	Gaps 14;

[illegible]

**RESULT 15**

(probable beta-glucosidase [imported] - Arabidopsis thaliana  
 C.Species: Arabidopsis thaliana (mouse-ear cress)  
 C.Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
 C.Accession: T02400; F84878  
 C.Round(s): S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,  
 submitted to the EMBL Data Library, May 1998  
 A.Description: Arabidopsis thaliana chromosome II BAC F411 genomic sequence.  
 A.Reference number: Z14667  
 A.Accession: T02400  
 A.Status: translated from GB/EMBL/DBDJ  
 A.Molecule type: DNA  
 A.Residues: 1-506 <ROU>  
 A.Cross-references: EMBL:AC004521; NID:g3128166; PIDN:MACT6091.1; PID:g3128187  
 A.Experimental source: cultivar Columbia  
 R.Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N  
 uss, D. H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-766, 1999  
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A.Reference number: A84420; MUID:20083487; PMID:10617197  
 A.Accession: F84878  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-506 <STO>  
 A.Cross-references: GB:AE002093; NID:g3128187; PIDN:AACT6091.1; GSPDB:GN00139  
 C.Genetics:  
 A.Gene: F411.26; AC2g44450  
 A.Map position: 2  
 A.Introns: 50/3; 74/1; 93/3; 119/1; 145/1; 174/2; 259/3; 370/1; 380/3; 415/1; 447/2  
 C.Superfamily: Agrobacterium beta-glucosidase

Query Match	35.0%;	Score 925.5;	DB 2;	Length 506;
Best Local Similarity	43.1%;	Pred. No. 2.8e-65;		

Matches 204; Conservative 71; Mismatches 169; Indels 29; Gaps 14;

```
QY      8 PNDPFWGFAATAAIOIEGAVKEGGRPSIMWTYCHLEPSR-TNGANGDVACDHYRYDEDF 66
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     37 PEDFIFGSATSATQVGEAGHEDGRGSIWDTFSEKYPEKIKDGSNGSVADNSYHLKEDV 96
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      67 DLLTKYGAKAYRPSLSNSRIIPLOGRLDPVNERGIEFYSKLIDALLRGGITTPWVTLXHMD 126
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     97 ALIHQIGFNAIRFESISWSRIILPRGNLKGGINOAGIDYNNLINELLSKGIKPPATMFHMD 156
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     127 LPGAHDRVGGMLNVEEVQLDPERYARLCFERFGDRVQNMWITINXPMWIAIYGATGSNA 186
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     157 TPQALEDAYGFGAGETVN-DFRDYADICFKNFGDRVKGMMTILNEPLTVVQOGYVAGVMA 215
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     187 PGRSS--INHGTEGNTATEPWLAKQAQIMSHARAVAVYSRDRPPSQQIGISLNGDY 244
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     216 PGRSKFTNCTDNGATEPIYVGHNLILSHGAAVQVREKYYKASQOGVGIALNAGWN 275
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     245 EPWDSNEPRDKEAERMEFHIGWFANPIFLKDYPESMKKQIGERLPALTPADFAILNA 304
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     276 LPY-TESPKDRLAARAAAFTEFDYFMEPLVTGK-YPVDVNNVKGRLPIFTAQOSKMLK- 332
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     305 GETDFYGMNYYTSQFARHLDPVPETD--YLGAIHHEQENKDGSPVGEESGLAMLRSCP 361
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     333 GSYDFIGINYSSTYAK--DVPCSTKDVMTMSDPCASVTGERDGVPIGPKAASDWLIIYP 390
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     362 DMFRKHLARVYGLYK-----PI-YITENGCPGPEENMTCEEAVNDPFRIRYFDSHDS 415
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     391 KGIRDLV-----LYAKYKFKDPVMVYITENG----RDEFSTNKIFLKDGDRIDYARHLEM 441
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     416 ISKAITQDGVVVKGFAMALLDNLWMSDYGPRPGVFTDYTT-LKRPKKA 467
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     442 VQDAISV-GANVKGFFAWSLDNFEWAMGYTVRFGLVYVDPKDGCKRYPKKA 493
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: March 25, 2004, 15:57:08  
Job time : 22 secs



XX Claim 18; Fig 2; 21pp; English.

XX The invention discloses a substantially purified BGL5 polypeptide with  
PS the biological activity of a beta-glucosidase. Also claimed is an  
CC isolated polynucleotide derived from a fungal source, which comprises a  
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,  
CC a Trichoderma reesei bgl5 nucleic acid coding sequence, or its  
CC sequence, an expression construct including the T.reesei bgl5 cDNA  
CC a recombinant host cell comprising a deletion or insertion or other  
CC alteration in the bgl5 gene which inactivates the gene and prevents BGL5  
CC polypeptide production, an antisense oligonucleotide, a detergent  
CC composition comprising the polypeptide and expressing a heterologous  
CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5  
CC is useful for producing ethanol, which involves contacting a biomass  
CC composition with an enzymatic composition comprising the polypeptide to  
CC yield a sugar solution, adding to the sugar solution a fermentative  
CC microorganism and culturing the fermentative microorganism under  
CC conditions sufficient to produce ethanol, where the biomass composition  
CC may be optionally pretreated. The method further involves the addition of  
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with  
CC a dilute acid. BGL5 is useful in wine making for enhancing the potential  
CC aroma of the finished wine product. The detergent is useful as softening  
CC agent and also useful for improving the feel of cotton fabrics and for  
CC degrading wood pulp into sugars. The sequence presented is the T. reesei  
CC BGL5 protein.

XX Sequence 484 AA;

Query Match 99.4%; Score 2629; DB 7; Length 484;

Best Local Similarity 99.8%; Pred. No. 7e-254;

Matches 483; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPESALNDPEWGFATAYOIEGAVKEGSGPSIMPTYCHLEPSRTKANGADYACDHYH 60  
DB 1 MPESALNDPEWGFATAYOIEGAVKEGSGPSIMPTYCHLEPSRTKANGADYACDHYH 60  
QY 61 RYDEDFDLITKYGAKAVRFSLSWSRIIPLGRLDPVNEEGIEFYSKIDALLRGITPMW 120  
DB 61 RYDEDFDLITKYGAKAVRFSLSWSRIIPLGRLDPVNEEGIEFYSKIDALLRGITPMW 120  
QY 121 TLVHMDLPQALHLDHYGGLNVEEYQVLDPERYARLCFEREGDGVONWITINXPMIOAIYGY 180  
DB 121 TLVHMDLPQALHLDHYGGLNVEEYQVLDPERYARLCFEREGDGVONWITINXPMIOAIYGY 180  
QY 181 ATGSNAPGRSSINHSSTEGNTATEPMLAGKAQIMSHARAVVYSRDRPSQKQIGISLN 240  
DB 181 ATGSNAPGRSSINHSSTEGNTATEPMLAGKAQIMSHARAVVYSRDRPSQKQIGISLN 240  
QY 241 GDYIEPMDSNPRDKAERMEFHIGWFANPIFLKDYEPESMKQGLERLPALTTPADFA 300  
DB 241 GDYIEPMDSNPRDKAERMEFHIGWFANPIFLKDYEPESMKQGLERLPALTTPADFA 300  
QY 301 ILNGETDFYGMNYYTQSFARHLDGPVPEVDYLCAIHEHQNKDGSVVGESGLAMTRSC 360  
DB 301 ILNGETDFYGMNYYTQSFARHLDGPVPEVDYLCAIHEHQNKDGSVVGESGLAMTRSC 360  
QY 361 PDMEFKHLARVYGLYKPRYITTEGNCPCPEENMTCEAVNDPRIRYFPSSHLDSTSKAI 420  
DB 361 PDMEFKHLARVYGLYKPRYITTEGNCPCPEENMTCEAVNDPRIRYFPSSHLDSTSKAI 420  
QY 421 TODGVVVKGYFAMALLDNLEWSDGYGRFGVFTDYTLTKRTPKSALVKDMPAARQRY 480  
DB 421 TODGVVVKGYFAMALLDNLEWSDGYGRFGVFTDYTLTKRTPKSALVKDMPAARQRY 480  
QY 481 KVAA 484  
DB 481 KVAA 484

RESULT 2  
AAG28767

ID AAG28767 standard; protein; 514 AA.  
XX  
AC AAG28767;  
AC  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34109.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
EN EPI033405-A2.  
PD  
FD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.

PR	8-JUN-1999	99US-0139461P
PR	8-JUN-1999	99US-0139462P
PR	18-JUN-1999	99US-0139463P
PR	18-JUN-1999	99US-0139464P
PR	18-JUN-1999	99US-0139763P
PR	18-JUN-1999	99US-0139765P
PR	21-JUN-1999	99US-0139691P
PR	22-JUN-1999	99US-0139699P
PR	23-JUN-1999	99US-0140353P
PR	23-JUN-1999	99US-0140354P
PR	24-JUN-1999	99US-0140659P
PR	24-JUN-1999	99US-01409821P
PR	28-JUN-1999	99US-0140981P
PR	29-JUN-1999	99US-0141287P
PR	30-JUN-1999	99US-0141287P
PR	01-JUL-1999	99US-0141842P
PR	01-JUL-1999	99US-0142154P
PR	02-JUL-1999	99US-0142055P
PR	02-JUL-1999	99US-0142390P
PR	08-JUL-1999	99US-0142803P
PR	08-JUL-1999	99US-0142920P
PR	12-JUL-1999	99US-0142977P
PR	13-JUL-1999	99US-0143542P
PR	14-JUL-1999	99US-0143624P
PR	15-JUL-1999	99US-0143763P
PR	15-JUL-1999	99US-0144005P
PR	16-JUL-1999	99US-0144085P
PR	16-JUL-1999	99US-0144086P
PR	19-JUL-1999	99US-0144325P
PR	19-JUL-1999	99US-0144332P
PR	19-JUL-1999	99US-0144333P
PR	19-JUL-1999	99US-0144334P
PR	19-JUL-1999	99US-0144335P
PR	20-JUL-1999	99US-0144352P
PR	20-JUL-1999	99US-0144632P
PR	20-JUL-1999	99US-0144684P
PR	21-JUL-1999	99US-0144814P
PR	21-JUL-1999	99US-0145086P
PR	21-JUL-1999	99US-0145088P
PR	21-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145087P
PR	22-JUL-1999	99US-0145089P
PR	22-JUL-1999	99US-0145192P
PR	22-JUL-1999	99US-0145145P
PR	23-JUL-1999	99US-0145218P
PR	23-JUL-1999	99US-0145224P
PR	26-JUL-1999	99US-0145276P
PR	26-JUL-1999	99US-0145913P
PR	27-JUL-1999	99US-0145918P
PR	27-JUL-1999	99US-0145919P
PR	28-JUL-1999	99US-0145951P
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147935P
PR	11-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148131P
PR	18-AUG-1999	99US-0149315P
PR	18-AUG-1999	99US-0149246P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	23-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149929P

PR	23-AUG-1999	99US-01499310P
PR	25-AUG-1999	99US-01505656P
PR	26-AUG-1999	99US-0150884P
PR	27-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151080P
PR	27-AUG-1999	99US-0151030P
PR	30-AUG-1999	99US-0151438P
PR	31-AUG-1999	99US-01519330P
PR	01-SEP-1999	99US-0152363P
PR	07-SEP-1999	99US-0153070P
PR	10-SEP-1999	99US-0153758P
PR	13-SEP-1999	99US-0154018P
PR	15-SEP-1999	99US-0154039P
PR	16-SEP-1999	99US-0154779P
PR	20-SEP-1999	99US-0155139P
PR	22-SEP-1999	99US-0155486P
PR	23-SEP-1999	99US-0155659P
PR	24-SEP-1999	99US-0155648P
PR	28-SEP-1999	99US-0156536P
PR	29-SEP-1999	99US-0157117P
PR	04-OCT-1999	99US-0157753P
PR	05-OCT-1999	99US-0157865P
PR	06-OCT-1999	99US-0158029P
PR	07-OCT-1999	99US-0158222P
PR	08-OCT-1999	99US-0158363P
PR	12-OCT-1999	99US-0159233P
PR	13-OCT-1999	99US-0159294P
PR	13-OCT-1999	99US-0159295P
PR	14-OCT-1999	99US-0159330P
PR	14-OCT-1999	99US-0159331P
PR	14-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159638P
PR	18-OCT-1999	99US-0159584P
PR	21-OCT-1999	99US-0160741P
PR	21-OCT-1999	99US-0160767P
PR	21-OCT-1999	99US-0160768P
PR	21-OCT-1999	99US-0160770P
PR	21-OCT-1999	99US-0160814P
PR	21-OCT-1999	99US-0160815P
PR	22-OCT-1999	99US-0160980P
PR	22-OCT-1999	99US-0160981P
PR	22-OCT-1999	99US-0160989P
PR	25-OCT-1999	99US-0161404P
PR	25-OCT-1999	99US-0161405P
PR	25-OCT-1999	99US-0161406P
PR	26-OCT-1999	99US-0161359P
PR	26-OCT-1999	99US-0161360P
PR	26-OCT-1999	99US-0161361P
PR	28-OCT-1999	99US-0161920P
PR	28-OCT-1999	99US-0161932P
PR	29-OCT-1999	99US-0162142P

Query Match	39.2%;	Score 1037;	DB 3;	Length 514;
Best Local Similarity	44.1%;	Pred. No. 2e-94;		
Matches 209;	Conservative 85;	Mismatches 156;	Indels 24;	Gaps 10;

```

Oy 6 ALPNDPMEGPEFATAYOJEGAVKEGGRPSIDYTCHEPSTNANGVACDHARVED 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 29 SFGPGFVFCGTAASSAFQHEGAVKAEGRPTIMDTSHFGKTTDPSNMDVANDOYHRRED 88
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 66 FDLITTKYGAARYPSLSWSRITPLGGRLPDVNEGIEFYSKLIDALLRRGITTPWTLYHM 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 89 VQLMKNMGMDAYRPSISWTIRIFPNNG-VGHINEGIDHYNNLINALAKGIEPVTLTYHM 146
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 126 DLPOALHDYRGYMLNVEEVOIDFERVYARLCPFERGDRVQNMITTNKPMIOAIYVATGSN 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 DLPOALHDRIYGLMNPQIIN-DFAAYAVEFCQRGRDVKMHMITNEHFTPALIGDYVGLQ 205
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 186 APGRSSINKKST-EGNTATEPWLAKAQIMSHARAVVYSRDRPBGOGIIGISLNGDY 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 206 AFGRTIILFKLTREGNSSTEPYIVGHNVILTHATVSDIYRKXKAKOGSLGIAFDWM 265  
QY 244 YEPMDNSEPROKEAREMERMEHFIQFANPIFLKQDPESSMKQGLCEBRI.PALTPADFALN 303  
Db 266 FEP-BSNTEDELEAQRADDFOLGWFLLPLMF-GYPSMSRSRVSLPVLFTGSSSLVX 323  
QY 304 AGETDFYGMNYYTSQFAHLLDGPVPEYDYLGAIHGHQNKQJ-----SPVGEESGL 354  
Db 324 -GSUDFVGINHTTYARN-----NATWLIGTLHDAVSDGTYLTPFKGLSTIGDRASS 377  
QY 355 AMLBSCPMEFRKHARVYGLYK-DYITTEGCGPCGSENMTCBEAVNDPFRIRYFDSHL 413  
Db 378 IWLTVLPGRMSRSLMYIKGRYGNPVPVFTTEGMDPNISILSRKDALDKARIKYHNDYL 437  
QY 414 DSISKAITQDGVVYGFAMALLDNLSEMSDGYGPRFGVTFPTY--TTLKRTPKS 466  
Db 438 SSLQASIKEDGCNNGVTFWMSLLDNWMAAGISSRFGLYFVDYRDLNLRKYPKDS 491

RESULT 3  
AAG28766  
ID AAG28766 standard; protein; 516 AA.  
XX  
AC AAG28766;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SHQ ID NO: 34108.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hydridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0131825P.  
PR 05-MAR-1999; 99US-0133180P.  
PR 09-MAR-1999; 99US-0133548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139482P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142820P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.



PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147302P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147260P
PR	06-AUG-1999	99US-0147303P
PR	06-AUG-1999	99US-0147416P
PR	06-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147935P
PR	10-AUG-1999	99US-0148171P
PR	11-AUG-1999	99US-0148319P
PR	12-AUG-1999	99US-0148434P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148684P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149475P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	20-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	25-AUG-1999	99US-0150864P
PR	27-AUG-1999	99US-0151085P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151303P
PR	31-AUG-1999	99US-0151438P
PR	01-SEP-1999	99US-0151920P
PR	07-SEP-1999	99US-0152363P
PR	10-SEP-1999	99US-0153070P
PR	24-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155659P
PR	28-SEP-1999	99US-0156458P
PR	28-SEP-1999	99US-0156566P
PR	04-OCT-1999	99US-0157117P
PR	06-OCT-1999	99US-0157753P
PR	06-OCT-1999	99US-0157865P
PR	07-OCT-1999	99US-0158029P
PR	08-OCT-1999	99US-0158232P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0159283P
PR	13-OCT-1999	99US-0159284P
PR	13-OCT-1999	99US-0159285P
PR	14-OCT-1999	99US-0159329P
PR	14-OCT-1999	99US-0159330P
PR	14-OCT-1999	99US-0159331P
PR	14-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159638P
PR	18-OCT-1999	99US-0159684P
PR	21-OCT-1999	99US-0160741P
PR	21-OCT-1999	99US-0160767P
PR	21-OCT-1999	99US-0160768P
PR	21-OCT-1999	99US-0160770P
PR	21-OCT-1999	99US-0160814P
PR	21-OCT-1999	99US-0160815P
PR	22-OCT-1999	99US-0160980P
PR	22-OCT-1999	99US-0160981P
PR	22-OCT-1999	99US-0160989P
PR	25-OCT-1999	99US-0161404P
PR	25-OCT-1999	99US-0161405P
PR	25-OCT-1999	99US-0161406P
PR	26-OCT-1999	99US-0161359P
PR	26-OCT-1999	99US-0161360P
PR	26-OCT-1999	99US-0161361P
PR	28-OCT-1999	99US-0161920P
PR	28-OCT-1999	99US-0161929P

Query Match	39.2%; Score 1037; DB 3; Length 516;
Best Local Similarity	44.1%; Pred. No. 2e-94;
Matches	209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;
QY	6 ALPNDFEWGFATAAYQLEGAVKEGGRPSIWDYVCHLEPSRTNGANGDVACDHYRRDED 65
DB	31 SFGKGFVFGTASAFQHEGAVKAEGRGPTTIWDTSFHTGKITDPSNADVAVDQYHRYEED 90
QY	66 FDLITKYGAKAYRPSLSMSRIIPICGRLDPNNEGIEFYSKLIDALLRGITTPWTLVYHW 125
DB	91 VQLKNNMGMDYRPSISWTRILFENG--VGHINEAGIDHYNLINALAKGIEPVYTLVHW 148
QY	126 DLPOALHRYGGMVNEVEVDFFERYALCFERFGDRVQNNITINXPMIOIYGYATGSN 185
DB	149 DLPOALHRYIGMLNPOLIN-DFAAYAEVCFQRGDRYKMHITNEPHITPAIOGIDVQLQ 207
QY	186 APGSSSINKHST--EGNTATEPWLAKGAQIMSHARAVALVYSRDFRPSQGOIGISLNDY 243
DB	208 APGRCTILFLKTCREGNSSTEPYIVGHNVILTHATVSDIYRKVKYAKOGSGSLIAFDVWM 267
QY	244 YEPDSDNEPRKKAERMERMEHIGFANPIFLKODYPESMKQIGERIPALTPRADFALN 303
DB	268 FEP-ESNKTEIDIEAQAQADPOLGMFLDPLMF--GDYPSMSRVSGRILPVEFTGSSSLVK 325
QY	304 AGEDTFEGMNYVTOSFARHLIDGPVPEITDYLGAIEHQENKDG-----SPVGEESGL 354
DB	326 -GSIDPFVGINHYYTYARN-----NATNLIQTLHDAVSDSGTYTLPPKGLSTIGDRASS 379
QY	355 AMLRSCPDMPFRKHLARVYGLYK-PYIYTEGCPGCEENMTCEBAVNDPFRIRYFDSHL 413
DB	380 IWLTVPRGMSLNMVYIKHRYGNDPVPFITTEGMDDPSILISRKDALDAKRIKYHNDYL 439
QY	414 DSIKAITQDSVVVYKGFANMLLNLENKSDSGYGRFGVTFPDY-TTLKRTPKKS 466
DB	440 SSLQASIKEDCNKYGFVWMLLDNMEWAAGYSRFGILYFVDYRDLNLRKRYPKDS 493
RESULT 4	
AG828765	
ID	AA828765 standard; protein; 523 AA.
AC	
AG828765;	
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 34107.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
OS	Arabidopsis thaliana.
XX	
XX	EP1033405-A2.
PV	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-00301439.
XX	
PR	25-FEB-1999; 99US-0121825P.
PR	05-MAR-1999; 99US-0123180P.
PR	09-MAR-1999; 99US-0123548P.
PR	23-MAR-1999; 99US-0125788P.
PR	25-MAR-1999; 99US-0126264P.
PR	29-MAR-1999; 99US-0126785P.
PR	01-APR-1999; 99US-0127462P.
PR	06-APR-1999; 99US-0128234P.
PR	08-APR-1999; 99US-0128714P.
PR	16-APR-1999; 99US-0129845P.
PR	19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 26-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-014256P.  
PR 14-MAY-1999; 99US-014218P.  
PR 14-MAY-1999; 99US-0142219P.  
PR 14-MAY-1999; 99US-014370P.  
PR 14-MAY-1999; 99US-014370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-014941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-013911P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 13-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-015753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.

PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 39.2%; Score 1037; DB 3; Length 523;

Best Local Similarity 44.1%; Pred. No. 2.1e-94;  
Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

QY 6 ALPNDPFWGATATAYQIEGAVKEGGRGPSIWDITYCHLEPRTNGANDVACDHYRYDED 65  
DB 38 SFRKGFVFGTASSAFQHEGAVKAEGRPTIWDTFSHFGKITPFSNADVAVDQYHRYEED 97  
QY 66 FDLITKYGAKAYRPSISMSRIIPLGRLDPVNEEGIEFYSKLDALLRGRITPMTLYHM 125  
DB 98 VOLMKMGMDAYRPSISMTNIPFNG--VGHINEAGIDHYNKLINALAKGIEPTVILYHW 155  
QY 126 DLPOALHRYGMLNVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAYGYATGSN 185  
DB 156 DLPOALHRYGMLNVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAYGYATGSN 214  
QY 186 AGRSSINRST--EGTATPEMLAGCAQIMSHARAVAVYSRDFRPSQKQIGISLNGDY 243  
DB 215 APRCCTILFPLTCREGNSSTEPYIVGHNVILTHATVSDIYKXKXKAKQGSGLGIAPFVW 274  
QY 244 YEPWDSNEPRDKAARMEFHIGWFANPIFLKXDYBESKKQUGRLPALTPADFAILN 303  
DB 275 FEP-ESNKTETDIEAQAQDFOLGFWLDPIMF--GDYPSMSRVSRLPVFTGSQSSLVK 332  
QY 304 AGETDFYGMNYYTSQPARHLDPVPETDYLGAIEHQENKDG-----SPVGEESGL 354  
DB 333 -GSLDFVGINHYTYYARN-----NATNLIGTLHDAVSDSGTYTLRPFKGLSTIGDPA 386  
QY 355 AWLRSCPMFRKLARVGLYGR--PIYTENGCCPGEENMTCEAVNDPRIRYFDSHL 413  
DB 387 IWLVIYPRGRSLMNYIKHRYGNPVPFITENGMDPSILSRKDALCKAKRIKYHHDYL 446  
QY 414 DSLSKAITODGVVVKYFAWALLDNLMSDYGGRPGVTFDYTT-LKRTPKSKS 466  
DB 447 SSLQASTIKEDGCVKGYFVMSLIDNMWEMAGYSSRFELVFDYRDNLKRYPKDS 500

## RESULT 5

AAB26813 standard; protein; 507 AA.

XX AAB26813;

XX 23-JAN-2001 (first entry)

XX Beta-primaverosidase mature protein amino acid sequence.

XX Beta-primaverosidase; beta-primeverose; tea aroma.

XX

OS Camellia sinensis.  
XX  
PN WO200052177-A1.  
XX  
PD 08-SEP-2000.  
XX  
PF 02-MAR-2000; 2000WO-JP001242.  
XX  
PR 04-MAR-1999; 99JP-00056299.  
XX  
PA (AMANO) AMANO PHARM KK.  
XX  
PI Sakata K. Mizutani M;  
XX  
DR WPI: 2000-572189/53.  
DR N-PSDB; AAA99407.  
XX  
XX beta-primaverosidase gene and variants obtained easily and cheaply by  
PT expression in *Escherichia coli* under conventional conditions, useful for  
PT production of tea aroma component beta-primeverose by its action on beta-  
PT primaveroside.

PS Example 4; Page 25-27; 30pp; Japanese.

CC This invention relates to a beta-primaverosidase gene and the protein it  
CC encodes. The invention includes a vector containing a DNA fragment of the  
CC beta-primaverosidase gene, a transformant containing the vector, and a  
CC process for producing beta-primaverosidase by culturing the transformant.  
CC The gene and its variants can be used to produce enzymes involved in the  
CC production of the tea aroma component beta-primeverose. The present  
CC sequence represents the beta-primaverosidase enzyme amino acid sequence  
CC of the invention

XX Sequence 507 AA;

QY 6 ALPNDPFWGATATAYQIEGAVKEGGRGPSIWDITYCHLEPRTNGANDVACDHYRYDE 64

DB 38 SFRKGFVFGTASSAFQHEGAVKAEGRPTIWDTFSHFGKITPFSNADVAVDQYHRYEED 97

QY 65 FDLITKYGAKAYRPSISMSRIIPLGRLDPVNEEGIEFYSKLDALLRGRITPMTLYHM 124

DB 98 DVAVLKFIFIGDGRMSISMARVLPKGLSGGVKKEGIAFYNNYINDLSGIGPFIIFH 157

QY 125 WDLPOALHRYGMLNVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAYGYATGS 184

DB 158 WDLPOALHRYGMLNVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAYGYATGS 216

QY 185 NARGR--SSINRSTEGTATPEMLAGCAQIMSHARAVAVYSRDFRPSQKQIGISLNGDY 243

DB 217 LAPGRSAPFAFPCPNNSGTEPIVTHNLLSHAAVVKYKEXYQAYQKQIGITLVTVW 276

QY 244 YEPWDSNEPRDKAARMEFHIGWFANPIFLKXDYBESKKQUGRLPALTPADFAILN 303

DB 277 MITY--SNSRADKQAQALDFMIGWIEPLSF--GEYPSMSRVSRLPVFTGSQSSLVK 333

QY 304 AGETDFYGMNYYTSQPARHLDPVPETDYLGAIEHQENKDG-----KDSPVGEESGLAWLRS 359

DB 334 KGSFDFLGINTYIANYV--LNVPTSNVSLYTTSDLSNQTAFRNCVAGRPVAFEM 391

QY 360 CPDMFRKLARVGLYGR--PIYTENGCCPGEENMTCEAVNDPRIRYFDSHLSISK 418

DB 392 YPKGLDVLVYTKKENDPIYITENG--MGDNNNVTEGIDPORYVFPYNOHLSLKN 449

QY 419 AITQDGVVVKYFAWALLDNLMSDYGGRPGVTFDYTT-LKRTPKSKSLVVK 471

DB 450 AIAA--GVKXKGYTMAFLDNFELSGTGRFGIIVYDFKDGRLKRYKSHLMFK 502

## RESULT 6

[illegible]

PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0156599P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159283P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 36.5%; Score 965; DB 3; Length 549;

Best Local Similarity 42.3%; Pred. No. 3.6e-87; Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDFEWCATAYOIEGAVKEGGRPSIMDTYCHLEPSP-TNGANGDVACDHYHYDE 64  
DB 7 SFRKGLFEGTASSSYQEGVNEGACQSWDHSNMFPHRISSSGCVAVADVFNHYKE 66  
QY 65 DPLLLTYKAKAVRFSISMSRIIPLGRLDPVNEEGIEFYSKLIDALLRGITPWTLYH 124  
DB 67 DIRKMKOINNDSEFSLIAMPRLVPGKRDGVSEEGIKFYNDVIDELLANEITPLVTIFH 126  
QY 125 WDIPOALHDYRGWLVNEEYQDPERYARLCFESEFGDRVONWITINXPMIOAIYVATGS 164  
DB 127 WDIPODLEDDEYGGFLS-EQITIDDRDYASLCFESEFGDRVSLWCTWNEPWSVAGYDTGR 185  
QY 185 NAGGRSS--INKISTEGNTATEPMLAGKAOIMSHARAVAVYSDRFRPSQGGQIGISLNGD 242

DB 186 KAPGRCSKYVNGASVAGMSGYEAYIVSHNMLLAHAEAVEVF-RKCDHIRNGQIGIAHNPL 244  
QY 243 YVEPMDSNEPRDDEAREMERMEFICGFANPIFLKQVPSMKQGLERLPALEPAFAL 302  
DB 245 WREYDPSDDVDGECNRANDFEMLGHQHPYAC-GDYETMKASVGDRLPSFTPEOSKUL 303  
QY 303 NAGETDFYGMNVTYSQFA--RHLDGPVP-ETDYLGAIHEHONEDGSPVGEESGLAWL 357  
DB 304 -IGSCDYGINVYSSLFVKSIXKIVDPTQPTWRDQ--GVDMKNTNIDGQIAKQGSSEWS 360  
QY 358 RSCPEDMFRKHLARVYLKQ-PIYITENGCPCEG----ENMTCEAVNDPFRIRYFDS 411  
DB 361 FTYPFTGLRNILKVKKTYGNPILITENGY---GEVAEQSQSLMYNPSIDTERLEYIEG 417  
QY 412 HLDISKAITODGVVVKYFAMALLNLNLEMSDGYGRFGVTFPDYTT-LKRPFKSALV 470  
DB 418 HIAIHQALHEDGVRVEGYVWSLLDNFEWNSGYVRYGLYIDYKGLRRYKMSALWL 477  
QY 471 KD 472  
DB 478 KE 479

RESULT 7  
AAG29188  
ID AAG29188 standard; protein; 560 AA.  
XX AC AAG29188;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34686.  
XX KM Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX OS Arabidopsis thaliana.  
XX PN EPI033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
XX PR 05-MAR-1999; 99US-0123180P.  
XX PR 09-MAR-1999; 99US-0123548P.  
XX PR 23-MAR-1999; 99US-0125788P.  
XX PR 25-MAR-1999; 99US-0126264P.  
XX PR 29-MAR-1999; 99US-0126785P.  
XX PR 01-APR-1999; 99US-0127462P.  
XX PR 06-APR-1999; 99US-0128234P.  
XX PR 08-APR-1999; 99US-0128714P.  
XX PR 16-APR-1999; 99US-0129845P.  
XX PR 21-APR-1999; 99US-0130077P.  
XX PR 21-APR-1999; 99US-0130449P.  
XX PR 23-APR-1999; 99US-0130510P.  
XX PR 23-APR-1999; 99US-0130891P.  
XX PR 28-APR-1999; 99US-0131449P.  
XX PR 30-APR-1999; 99US-0132048P.  
XX PR 04-MAY-1999; 99US-0132407P.  
XX PR 04-MAY-1999; 99US-0132484P.  
XX PR 05-MAY-1999; 99US-0132485P.  
XX PR 06-MAY-1999; 99US-0132486P.  
XX PR 07-MAY-1999; 99US-0132487P.  
XX PR 07-MAY-1999; 99US-0132863P.  
XX PR 11-MAY-1999; 99US-0134256P.  
XX PR 14-MAY-1999; 99US-0134218P.  
XX PR 14-MAY-1999; 99US-0134219P.  
XX PR 14-MAY-1999; 99US-0134221P.  
XX PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145221P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147329P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 25-AUG-1999; 99US-0150884P.  
PR 26-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 27-AUG-1999; 99US-0151303P.  
PR 30-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 13-OCT-1999; 99US-0159325P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 18-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 21-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.

PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 36.5%; Score 965; DB 3; Length 560;  
Best Local Similarity 42.3%; Pred. No. 3.7e-87;  
Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDEMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64  
DB 18 SFGKGLFGRASSSYQYEGAVNEGARGOSVWDHFSNRFPHRISDSSDGNVAVDYHRYKE 77  
QY 65 DFDLTKYGAAYRSLISMSRIIPLGRLDPVNEEGIEFYSKIDALLRRGITPWTLYH 124  
DB 78 DIRKMDINMDSFRLSIAMPRLVPGKRDGRVSEEGIKFYNDVIDELLANEITPLVTIFH 137  
QY 125 WDLPOALHRYGGMVNEEVOUDFERYARLCFERFGDRVONMTTINXPMIOATGYATGS 184  
DB 138 WDIPOULEDEYGGFLS-EQIIDDYDASLCFERFGDRVSLMCTMNEPWSVAGYDTGR 196  
QY 185 NABGRSS--INKSTEGNTATEPMLAGKAOIMSHARAVYSRDPSPQKQIGISLNGD 242  
DB 197 KAGRCRSKYVNGASVAGMSGYEAYIVSHNMLLAHAEVEVF-RKCHIKKQIGIAHNPL 255  
QY 243 YPEPMDNEPRDKEAERMEFIHGFANDIFLKQYPSMKKQIGERLPAITPADFAIL 302  
DB 256 WYEPYDPSDDVEGCNRAMDFMLGWHQHPAC-GDYPTMKKSVGDRLPSPFPEQSKKL 314  
QY 303 NAEETDYGNNYTSQFA---RHLDGVP--ETDYLGAIHEHENDGSGVSGESGLAWL 357  
DB 315 -IGSCDYVGINYSLSFVKSIRKVDPTQPTWRDQ--GVMMKTNIDGKQIAKQSGSEWS 371  
QY 358 RSCPDWFRKHLARVGLYK-PYITENGCPCE-----ENMTCEAVNDPFRIRYFDS 411  
DB 372 FTYPTGLRNILKTKYKTYGNPILITNGY---GEVAEQSOSLYMNPISIDTERLEIIEG 428  
QY 412 HLDISKAITQDGVVVKGYFAMALLDNLEWSDGYGPRFGVTFDYTT-LKRTPKKSALVL 470  
DB 429 HIAHIOAIHEDGVRVEGYVWMSLDNFEWNSGYGVAYIDYKDGRLRRYPMKSALWL 488  
QY 471 KD 472  
DB 489 KE 490

RESULT 8  
ADC03121  
ID ADC03121 standard; protein; 560 AA.  
XX  
AC ADC03121;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
XX A. thaliana family 1 glycoside hydrolase PEN2.  
XX beta-glucosidase; antibacterial; fungicide; plant protection; plant;  
KM PEN2; family 1 glycoside hydrolase.  
XX  
OS Arabidopsis thaliana.  
XX  
XX WO2003074688-A2.  
XX  
XX 12-SEP-2003.  
XX  
XX 06-MAR-2003; 2003WO-EP002315.  
XX  
XX 06-MAR-2002; 2002EP-00004400.  
XX  
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX  
PI Iapka V, Scheel D, Schulze-Lefert P, Rosahl S, Landtag J;  
XX WPI; 2003-712890/67.  
DR N-PSDB; ADC03120.  
XX  
XX New polynucleotide encoding a beta glucosidase plant protein, useful for  
PT the preparation of a plant protection composition, and/or for  
PT establishing or enhancing pathogen resistance in plants.  
XX  
PS Claim 1; SEQ ID NO 2, 89pp; English.  
XX  
XX The invention relates to a novel polynucleotide encoding a beta-  
CC glucosidase protein. A polypeptide of the invention has antibacterial and  
CC fungicide activity. The identified compound or hydrolysis product of the  
CC invention is useful for preparing a plant protection composition. The  
CC present sequence represents the Arabidopsis thaliana PEN2 protein.  
SQ Sequence 560 AA;

Query Match 36.5%; Score 965; DB 7; Length 560;  
Best Local Similarity 42.3%; Pred. No. 3.7e-87;  
Matches 204; Conservative 85; Mismatches 169; Indels 24; Gaps 13;

QY 6 ALPNDEMGFATAAYOIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGVACDHYHRDE 64  
DB 18 SFGKGLFGRASSSYQYEGAVNEGARGOSVWDHFSNRFPHRISDSSDGNVAVDYHRYKE 77  
QY 65 DFDLTKYGAAYRSLISMSRIIPLGRLDPVNEEGIEFYSKIDALLRRGITPWTLYH 124  
DB 78 DIRKMDINMDSFRLSIAMPRLVPGKRDGRVSEEGIKFYNDVIDELLANEITPLVTIFH 137  
QY 125 WDLPOALHRYGGMVNEEVOUDFERYARLCFERFGDRVONMTTINXPMIOATGYATGS 184  
DB 138 WDIPOULEDEYGGFLS-EQIIDDYDASLCFERFGDRVSLMCTMNEPWSVAGYDTGR 196  
QY 185 NABGRSS--INKSTEGNTATEPMLAGKAOIMSHARAVYSRDPSPQKQIGISLNGD 242  
DB 197 KAGRCRSKYVNGASVAGMSGYEAYIVSHNMLLAHAEVEVF-RKCHIKKQIGIAHNPL 255  
QY 243 YPEPMDNEPRDKEAERMEFIHGFANDIFLKQYPSMKKQIGERLPAITPADFAIL 302  
DB 256 WYEPYDPSDDVEGCNRAMDFMLGWHQHPAC-GDYPTMKKSVGDRLPSPFPEQSKKL 314  
QY 303 NAEETDYGNNYTSQFA---RHLDGVP--ETDYLGAIHEHENDGSGVSGESGLAWL 357  
DB 315 -IGSCDYVGINYSLSFVKSIRKVDPTQPTWRDQ--GVMMKTNIDGKQIAKQSGSEWS 371  
QY 358 RSCPDWFRKHLARVGLYK-PYITENGCPCE-----ENMTCEAVNDPFRIRYFDS 411  
DB 372 FTYPTGLRNILKTKYKTYGNPILITNGY---GEVAEQSOSLYMNPISIDTERLEIIEG 428  
QY 412 HLDISKAITQDGVVVKGYFAMALLDNLEWSDGYGPRFGVTFDYTT-LKRTPKKSALVL 470  
DB 429 HIAHIOAIHEDGVRVEGYVWMSLDNFEWNSGYGVAYIDYKDGRLRRYPMKSALWL 488  
QY 471 KD 472  
DB 489 KE 490

RESULT 9  
ABP71306  
ID ABP71306 standard; protein; 545 AA.  
XX  
AC ABP71306;  
XX  
DT 28-APR-2003 (first entry)  
XX  
XX Prunus serotina prunasin hydrolase.  
DE Prunasin hydrolase; transesterification; vascular tissue; gene expression;  
XX abiotic stress; biotic stress.  
KM





```
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142380P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0143297P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 15-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144355P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 22-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151030P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159684P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160816P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161559P.
PR 26-OCT-1999; 99US-0161607P.
PR 26-OCT-1999; 99US-0161610P.
PR 26-OCT-1999; 99US-0161616P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 35.5%; Score 939; DB 3; Length 512;
Best Local Similarity 41.4%; Pred. No. 1.3e-84;
Matches 197; Conservative 84; Mismatches 175; Indels 20; Gaps 11;

QY 6 ALPNDFEMGFATATAYOIEGAVNKGGRGPGSIWDTYCHLBESRRTGANGADVACDHYHYKDED 65
DB 43 SFPKGFVFGFATSAAYOVEGRTHODGRGPGSIWDAFVFI PGKIAKNAATAEITVDQYHRYKED 102
QY 66 FDLTLTYGAKAYFRFSWSRRIIPLGRLDPVNEGIEFYSKLDLALRGITPMTLYH 125
DB 103 VDLAKKLINPDATYRFSISWSRIIPEGS- GKVMKGVATYRRLIDYVQKITYPAULYH 160
QY 126 DLFOALHRYGGLNTEVEVQDPERYARLCFERFGDRVQNMITINXPMIQAIYVATGSN 185
```

Db 161 DLPLALENKKYKLLG-RQVYKDFADYAEFCYKTFGDRVKNNMTFNENRVVAAALGYDNGIF 219  
QY 186 APCR-SSINHGSTEPGNTATEPMLDAKQIMSHARAVAVSRPFRSQKQIGISINDGY 244  
220 AGRCKKAGCTGCTEGSATEPPIVTHHLILALAAVORRYKRYQAKQKRGVGLILDFFWY 279  
QY 245 EPMDSNEPPDKAEARMEFHIGWPAFPIFLKKDYPSKQKOLGRLLPALTFPADPAIINA 304  
280 EPL-TRSKADNLAAQFARDPHIGWFIHPL-VYGEVPTKQNIKVKERLPFTKEKVMK- 336  
QY 305 GETDFGKMYYSQSFARHLDGVPEDDYIGALHENOEN-----KDGSPVGEESGLAMIR 358  
Db 337 GSIDFPGINQVYTTY--NSEHPFTTKPKDGLYQDDMNVFEGFALGKPIGRAYSSWLY 393  
QY 359 SCPDMFRKILARVYGLYKRP-IYITENGCPGCEENMTCEEAANDPFRIRYFDSHLSIS 417  
Db 394 NVPWGVYKALMYMKERYGNPTMILSENGWDDG--NVTLAQGLHTTITIKYKQYLTMLK 451  
QY 418 KAITDDGVVVKGYFPMALIDNLEMSDYGPRPGVTFTYTTIKRTPKKSALVLKDM 473  
452 KA-RDDGANVGVYFAMSLIDNPEMLSGYTSRFGIYVVDYKTLKRYPKMSAQWFKQL 506  
Db

RESULT 11

AA614174

ID AA614174 standard; protein; 521 AA.

AA614174;

17-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 13936.

KM Protein identification; signal transduction pathway; metabolic pathway;  
KW hydridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

OS Arabidopsis thaliana.

XX PN EPI033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138647P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.

PR	08-JUN-1989	99US-0145591
PR	02-AUG-1999	99US-0146386P
PR	02-AUG-1999	99US-0146388P
PR	02-AUG-1999	99US-0146389P
PR	03-AUG-1999	99US-0147038P
PR	04-AUG-1999	99US-0147204P
PR	04-AUG-1999	99US-0147202P
PR	05-AUG-1999	99US-0147192P
PR	05-AUG-1999	99US-0147263P
PR	06-AUG-1999	99US-0147470P
PR	06-AUG-1999	99US-0147416P
PR	09-AUG-1999	99US-0147493P
PR	09-AUG-1999	99US-0147955P
PR	10-AUG-1999	99US-0148119P
PR	11-AUG-1999	99US-0148311P
PR	12-AUG-1999	99US-0148341P
PR	13-AUG-1999	99US-0148565P
PR	13-AUG-1999	99US-0148664P
PR	16-AUG-1999	99US-0149368P
PR	17-AUG-1999	99US-0149115P
PR	18-AUG-1999	99US-0149426P
PR	20-AUG-1999	99US-0149722P
PR	22-AUG-1999	99US-0149723P
PR	20-AUG-1999	99US-0149929P
PR	23-AUG-1999	99US-0149902P
PR	23-AUG-1999	99US-0149930P
PR	25-AUG-1999	99US-0150566P
PR	26-AUG-1999	99US-0150884P
PR	26-AUG-1999	99US-0151065P
PR	27-AUG-1999	99US-0151066P
PR	27-AUG-1999	99US-0151080P
PR	30-AUG-1999	99US-0151103P
PR	31-AUG-1999	99US-0151438P
PR	01-SEP-1999	99US-0151930P
PR	01-SEP-1999	99US-0151779P
PR	07-SEP-1999	99US-0152263P
PR	10-SEP-1999	99US-0153707P
PR	13-SEP-1999	99US-0153758P
PR	15-SEP-1999	99US-0154040P
PR	16-SEP-1999	99US-0154031P
PR	20-SEP-1999	99US-0154779P
PR	22-SEP-1999	99US-0155139P
PR	23-SEP-1999	99US-0155486P
PR	24-SEP-1999	99US-0155659P
PR	28-SEP-1999	99US-0156458P
PR	29-SEP-1999	99US-0156566P
PR	04-OCT-1999	99US-0157117P
PR	05-OCT-1999	99US-0157573P
PR	06-OCT-1999	99US-0157865P
PR	07-OCT-1999	99US-0158029P
PR	08-OCT-1999	99US-0158232P
PR	12-OCT-1999	99US-0158369P
PR	13-OCT-1999	99US-0159224P
PR	13-OCT-1999	99US-0159284P
PR	13-OCT-1999	99US-0159285P
PR	14-OCT-1999	99US-0159329P
PR	14-OCT-1999	99US-0159330P
PR	21-OCT-1999	99US-0160767P
PR	21-OCT-1999	99US-0160768P
PR	21-OCT-1999	99US-0159637P
PR	14-OCT-1999	99US-0159638P
PR	14-OCT-1999	99US-0159688P
PR	21-OCT-1999	99US-0160815P
PR	22-OCT-1999	99US-0160980P
PR	22-OCT-1999	99US-0160981P
PR	22-OCT-1999	99US-0160989P
PR	25-OCT-1999	99US-0161404P
PR	25-OCT-1999	99US-0161405P
PR	26-OCT-1999	99US-0161406P
PR	26-OCT-1999	99US-0161359P

PR 26-OCT-1999; 99US-0161360P  
PR 26-OCT-1999; 99US-0161361P  
PR 28-OCT-1999; 99US-0161920P  
PR 28-OCT-1999; 99US-0161992P  
PR 28-OCT-1999; 99US-0161993P  
PR 29-OCT-1999; 99US-0162142P

Query Match	35.5%	Score 939	DB 3	Length 521
Best Local Similarity	41.4%	Pred. No.	1.3e-84	
Best Match 197	Conservative	84	Mismatches 175	Indels 20
				Gaps 11

```

QY      6 ALPNDEMGFATAAQIEGAVGEGRGPSIMDTYCHLESRTNGANGDVACDHYHRDED 65
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      52 SFPKGFVFGTATSAYQVEGETHQDGRGSPIMDAFVKIPKIAKNAITAEITVDQYHRYKED 111

```

```
Qy      66 FDDLTKYGAAAYRRFSLSMSRKIPPLGRLDIPNNEEGIEFYFSKIDALLRGITFWVTLYHW 129
```

---

```
Db     112 VDLMKKLFPDAYRFSISMSRFPEGS--GKNMWMGAYYNNRLIDYMQIGITPYANLYHY 165
```

```
QY      126 DLPOLHDRGGGWLVEEVQDFFERARLCERFGBRVQNMTITXKMWIOALGYATSSN 185
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      170 DLPLALENKKGGLG-RQVVKDPAADAEFCYKTFGBRVKNMNTFNEBPRVAALGYDNCF 228
```

```
Qy      186 APGR-SSINXHESTENTATEWMLAKQAQIMSHARAAVYSDRFRPSQGQIGISLNDYY 244
        |||| : |||: |||: |::| | : : |||: || : : |
Db      229 AFGKSKAFNGCTEGNSATERYIVTHLLLAHAAAQRYRKYYQAKXGKGVGLLDFTVMY 288
```

```

Qy      245 EPWDSNEPRKEAAERMEPHIGWFPANPIFLKDDYBESMKKOLGERLPALTTPADFALINA 304
      || : || : || || : : || : : || : : || : : || : : || : : || : :
Db      289 EPL-TRSKADNLAAQRARDFHIGWFIHPL-VYGEYPTKMNIIVKERLPKFTKEVKWKV- 345

```

```
QY      305 GETDEPGMNYTTSOFARHLDSGPVEPDYLGAIHENGEN-----KSSPVGEESGLAFLR 355
          |||::||::| | | | | | | | | | | | | | | | | | | | | | | | | |
Db      346 GSIDFVGINGDTTTY--MSERPHPTTKPKDYGQQDMNVERGFAGLKGPIGPRAYSSWLY 402
```

QY 359 SCPDMFRKHLARVYGLYKPK-IYITENGCCPCGEEANNTCEEAVNDPRIRIRYFDSHLDSIS 417

DB 403 NVPMGMYYKALMYMKERYGNPTMLSENGMDPG--NVTLAGGLHDTTRIKYKYDYLTLNLIK 466

```

QY.      418 KAITQGGVVKVGFAMALLDNLENSDGYGRFGVFETDYYTLKRTPKKSALVLKDM 473
          || || || || || || || || || || || || || || || || || || || ||
Db        461 KA-RDDGANVVGFFAMSLLDNFEMLSGYNSRFGIVVDYKTLKRYPKNSAQWFQQL 515

```

RESULT 12  
ABP99404

AC ABP99404;  
XX

XX  
DE Arabidopsis thaliana polypeptide SEQ ID NO 93.  
XX

KW herbicide-resistant.  
XX  
OS *Arabidopsis thaliana*.

PN	WO200266660-A2
XX	
PD	29-AUG-2002.

PF	13-FEB-2002; 2002WO-EP001466
XX	
PR	16-FEB-2001; 2001DE-01007843

XX  
PA (META-) METANOMICS GMBH & CO KGAA  
XX

XX  
DR WPI; 2002-674953/72.

PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 08-JUL-1999; 99US-0142380P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149375P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 23-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.

PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155569P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160771P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 22-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161820P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 35.0%; Score 925.5; DB 3; Length 506;  
Best Local Similarity 43.1%; Pred. No. 2,9e-83;  
Matches 204; Conservative 71; Mismatches 169; Indels 29; Gaps 14;

QY 8 PNDPEMGPTAAVQIGAVGEGGRGSPIMDTYCHLEPSR-TNGANGDVACDHYRXYDEDF 66  
DB 37 PEDPFGSATSAYQVGGCAHEDGRGSIIMDTSEKTPBEKTKDSSNGSVADNSHLYKEDV 96  
QY 67 DLLTKYGAAYRPSLSWSRIIPVGRGLDPVNEBGEIEFYSLKIDALLRGITTPWVTLVHMD 126  
DB 97 ALHQIGFNAVYRISMSRILPRGNLKGGINQIDYNNLISLHLLSKIKPRATMHHMD 156  
QY 127 LPQALHDRYGMNLNVEVQIDFERIYARLCEPREGDGVQWITINXWICAIYGAAGSNA 186  
DB 157 TPQALBDAYGGFGGAGIYN-DFRDYADICFKNGGDEVKMMWITNEPLTVVQGGYVAVGWMA 215  
QY 187 PGRSS--INKSTEGTATPSPMLAGKAQIMSHARAVAYSRDPRPSOKGIGISLNDGY 244  
DB 216 PGRCSKFTPNCTDNGATTEPIYVGNLISHAAYQVYREKTKAQQGQGVGIALNAGWN 275  
QY 245 EPWDSNEPRDKAEARMEFHIGWFPANPIFLKKDYESNMKKQIGERI,PAITPADFAILNA 304  
DB 276 LPY-TESPQRDLAABAMAFTFDYFMEPIYVTK-YVDMVNNVKGRLPIFTAQOSKWLK- 332  
QY 305 GETDPFGMYTYTSQFARHLDGPYETD---YLGAIEHQNKGSGVGEBSGLAMLRSCP 361  
DB 333 GSYDFLGINYSYSTYAK--DVPCSTYDVTWTFSDPCASVTGERGVBIPGKAASDWLLIYP 390



PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136332P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139464P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140635P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142330P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145267P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.

PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 34.9%; Score 923; DB 3; Length 507;  
Best Local Similarity 41.9%; Pred. No. 5,2e-83;  
Matches 205; Conservative 75; Mismatches 173; Indels 36; Gaps 16;

QY 8 PNDPEMGFATAAYOIEGAVKGGRGSPSTMDPYCHLEPSR-INGANGDVACHYHRYDEDF 66  
DB 37 PKDPIFGAATISAYOIEGAHEDRGSPSTMDPSEKYPKIKGTNGSIASDSYHLYKEDV 96  
QY 67 DLATKYGAKAYRFSLSMSRIIPLGRLDPVNEBEGIEFYSKLIDALLRGTPTWVTLVHMD 126  
DB 97 GLHQIGFGAVRFSISWSRIILPRGNLKGGINAGIDYNNILNELSKIGKPFATIFHMD 156  
QY 127 LPQALHDPYGGMLNVEEYQLDPERYARLCFERFGDRVQNMITINXPMIQALYGYATGSNA 186  
DB 157 TPQSLIEDYVGGFPGAELYN-DFRDYADICFKNFGDRVGMMTLNEPLTVVQQGYVAGYMA 215  
QY 187 PGRSS--INKSTEGNTATEBPMLAGKQIMSHARAVNYSRDRPSOKGIGISLNGDY 244  
DB 216 PGRCSKFTNPNCTAGNGATEPYIVGHNLILAHGNAKVYREKYSQKQGVGIALNAGWN 275  
QY 245 EPWDSNEPRDEAAERMEFHIQWFPANFIPLKDYPSMKKOLGE-RLPALTPADFAILN 303  
DB 276 LPY-TESAEEDRLAARAAAFEDYFMEPLVTGK-YPVDVNNVAVDGRLPFTTAKQSMKLK 333  
QY 304 AGETDPYGMNYTQFARHLDGPVE-----TDYLAHIEHQENKXGSPVGEESGLAML 357  
DB 334 -GSTDFIGINITYSSYAK--DVPCSSENVTLFSDPCASVTGERE---GVPIGPKAASDWL 387  
QY 358 RSCPDMEKXHLARYGLYK-----PI-YITENGCPCEBENMTCEAVNDPPRIRYFDS 411  
DB 388 LIYFKGIRDL-----LYAKYKFKDPVWYITENG---RDEASTGKIDLKDSERIDYIAQ 438  
QY 412 HLDISKAITDGVVVGKGFAMALLDNLEWSDGYGPRFGVTFDYYT-LKRTPKKSAVL 470  
DB 439 HLKNVODALIS-IGANVVGFFAWSLDNFEMATGYSVRFGVYVDFNDGRKRYPKKSAXWF 497  
QY 471 KDMFAROR 479  
DB 498 RKLISEKKR 506

Search completed: March 25, 2004, 15:55:08  
Job time : 62 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 25, 2004, 15:54:02 ; Search time 22 Seconds  
(Without alignments) 1135.771 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644

Sequence: 1 MESLALPDEFWGFATAY.....KSLVTKDMPARQVVA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/PCFUS COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	983	37.2	551	3	US-08-688-988-37
2	981.5	37.1	507	4	US-09-914-841A-15
3	976	36.9	509	3	US-08-688-988-40
4	964.5	36.5	531	3	US-09-306-593-9
5	924.5	35.0	531	3	US-08-688-988-39
6	923.5	34.9	657	3	US-09-306-593-2
7	910	34.4	563	2	US-09-122-230-7
8	902	34.1	563	4	US-09-431-470-4
9	896	33.9	531	3	US-08-688-988-38
10	880.5	33.3	459	4	US-08-914-375C-72
11	879	33.2	566	4	US-09-431-470-2
12	863.5	32.7	524	3	US-08-688-988-34
13	861	32.6	429	4	US-09-914-841A-1
14	858.5	32.5	476	4	US-08-914-375C-69
15	858.5	32.5	548	3	US-08-688-988-41
16	857.5	32.4	470	4	US-08-914-375C-73
17	856	32.4	528	3	US-08-688-988-6
18	854	32.3	539	3	US-08-688-988-42
19	849.5	32.1	540	3	US-08-688-988-43
20	845	32.0	473	4	US-08-914-375C-71
21	842	31.8	471	4	US-08-914-375C-70
22	840.5	31.8	450	3	US-09-306-593-13
23	839	31.7	446	3	US-09-306-593-11
24	831	31.4	525	3	US-08-688-988-10
25	826	31.2	438	4	US-09-134-078-16
26	810	30.6	514	3	US-08-688-988-35
27	807	30.5	469	3	US-09-306-593-8

28	802	30.3	441	4	US-08-914-375C-66	Sequence 66, Appl
29	802	30.3	449	3	US-09-306-593-10	Sequence 10, Appl
30	791	29.9	439	4	US-08-914-375C-67	Sequence 67, Appl
31	790	29.9	425	3	US-08-688-988-36	Sequence 36, Appl
32	789	29.8	437	4	US-08-914-375C-65	Sequence 65, Appl
33	789	29.8	449	4	US-08-914-375C-68	Sequence 68, Appl
34	749	28.3	448	3	US-09-306-593-12	Sequence 12, Appl
35	747	28.3	441	4	US-08-914-375C-63	Sequence 63, Appl
36	745.5	28.2	356	4	US-08-914-375C-74	Sequence 74, Appl
37	726.5	27.5	450	4	US-08-914-375C-64	Sequence 64, Appl
38	718	27.2	1014	4	US-09-344-510B-3	Sequence 3, Appl
39	714	27.0	550	4	US-09-344-510B-4	Sequence 4, Appl
40	710	26.9	549	4	US-09-344-510B-2	Sequence 2, Appl
41	710	26.9	1012	4	US-09-344-510B-1	Sequence 1, Appl
42	708.5	26.8	437	3	US-08-688-988-8	Sequence 8, Appl
43	702.5	26.6	478	4	US-09-107-532A-4922	Sequence 4922, Ap
44	699	26.4	488	4	US-09-489-039A-9557	Sequence 9557, Ap
45	690.5	26.1	485	4	US-09-134-000C-6295	Sequence 6295, Ap

## ALIGNMENTS

RESULT 1									
US-08-688-988-37									
; Sequence 37, Application US/0868988B									
; Patent No. 6096545									
; GENERAL INFORMATION:									
; APPLICANT: Lefebvre, Daniel D.									
; APPLICANT: Malboobi, Mohammad A.									
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS									
; FILE REFERENCE: PPL96-03									
; CURRENT APPLICATION NUMBER: US/08/688,988B									
; CURRENT FILING DATE: 1996-07-31									
; NUMBER OF SEQ ID NOS: 48									
; SOFTWARE: FASTSEQ for Windows Version 3.0									
; SEQ ID NO 37									
; LENGTH: 551									
; TYPE: PRT									
; ORGANISM: Prunus serotina									
; US-08-688-988-37									
Query Match									
Best Local Similarity 45.3%; Pred. No. 5.3e-97;									
Matches 213; Conservative 70; Mismatches 171; Indels 16; Gaps 11;									
QY	6	ALPNDEWGFATAAYQIEGAVKEGGRGPSIMDTYCHLEPER-TNGANGDVACDHYRYDE	64						
DB	42	ALBPGFIFGTASAAYQFEGAKEDGRGPSIMDTYTHNHSERIKDGSNGDAVDQYHRYKE	101						
QY	65	DPDLTKYKGAARPSLSWSRITPLGRLDPVNEEGIEFYSKLDLILRRGITPWTLYH	124						
DB	102	DVIMKMGGDARFSLISWSRVLPNGVSGVNEDEGKFYNNILINELRGLRPPVITYH	161						
QY	125	WDLPQALHRYGGWLWVEEVDLDFERYARLCFERFGDGVQNMWITINXPMQAIYVATGS	184						
DB	162	WDLPQALHRYGGWLWVEEVDLDFERYARLCFERFGDGVQNMWITINXPMQAIYVATGS	220						
QY	185	NAPGRSSIT-NKSTEGNTATPEWLAGKQIOMSHARAAVAVSDFRPSQKQIGISLNGD	242						
DB	221	HAGRCGSAWQKLNCTGNSATPEYLVTHQLHAAAVKLYKDEYQASQNGLIGITLVSP	280						
QY	243	YVEPWSNPERDLEARRMEFHIGFANPIFLKQYPSRMKQGLRPLPADRAIL	302						
DB	281	WEPD-ASEADEDINAAFRSLDFFGWPMDBL-TNGVPHLMRSIVGRLLNFTBOSKLL	338						
QY	303	NAGETDFYGNMYTSQFARHLDPVETDYLGAIHHEQENK---DGSVGEESGLAMLR	358						
DB	339	K-GSPDGIQNTYTTTRVAS--NAPKITSVAASITTPQVVAATLKLGVPIGMAASGLW	395						
QY	359	SCDPMFRKHLARVYGLYKGP-IYITENGCPCEPEENNTCEAVNDPRIRIYFSSHLSIS	417						
DB	396	VYKGIHDLVLYTKEXYNDPLIYITENGVDPEFNDPKLSMEELKDTNRIDFYVRLCYLQ	455						

```
QY      418 KAITODGVVKGYPFAMALIDNLEMSDYGPRFGVTFTDYT-TLKRPKKS 466
      456 AAI-KKGSVKYKGYFAMSPLDNFEMDAGYVTRPOINVDINDILKRKSKLS 504

RESULT 2
US-09-914-841A-15
; Sequence 15, Application US/09914841A
; Patent No. 6645750
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme Inc.
; TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
; FILE REFERENCE: Q66072
; CURRENT APPLICATION NUMBER: US/09/914,841A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: P. Hei. 11-056299
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/JP00/01242
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 15
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Camelilia var. sinensis
US-09-914-841A-15

Query Match
Best Local Similarity 37.1%; Score 981.5; DB 4; Length 507;
Matches 206; Conservative 82; Mismatches 169; Indels 17; Gaps 12;

QY      6 ALPNDPENGATTAATVQIEGAVKEGGSGPSIMDTYCHLEPSR-TNGANGDVA CDHYHRDE 64
      38 SPDFGVFGAASASAYQFEGAKEGKGKGNIMDTFTHEPPKISNGSTGDVADDFYHRKE 97
      65 PDDLTKYGAAYRPSLSMSRIIPLGRLDPVNEGIEFFYSKLDLIRGITTWVLYH 124
      98 DVKVLKFTGLDGFNRSISWAVLPRGKLSGGVNEGLAFYNNVINDLSKGIOPFIYIFH 157
      125 WDLPGALHDYRGWLVNEEVQIDFERVYARLCEFERGDRVQNMITINXPMIOAIYATGS 184
      158 WDLPGALDEYGGFLSPHIV-DPRDFALCFKRGVDVKMITMNEPMSYSGYDAGL 216
      185 NAPER-SSINKHSTEGNTATEPWLAKQAQIMSHARAVYSRDPRPSQKQIGISLNGDY 243
      217 LAPGCSAFMAFCPRKNSGTEPYIVTNHLLSHAAVLYLEKYOAYQKQIGITLVYVW 276

QY      244 YEPMDNPRDKEAEERMEFHIGWFANPIFLKKDVPESMKKQIGERLPALTPADPAIIN 303
      277 MIPY-SNKAKDKDAQRALDPMYGFIEPLSF-GEYPSMKRLVKGKRLPFT-KEQAMLV 333
      304 AGETDFYMYNTYSQFARHLDPVPEYDYLGAIEHOEN---KDGSPVGEESGLAMLR 359
      334 KGSPEFLGNIYIANYV--LNVPTSNSVNLSTYDLSNQAIFNGLVAGIPGVPAFEM 391
      360 CPDMFRKHLAVYGLYGP-IYITENGCPCGGEENMCEAVNDPPRIKRFDSHLSISK 418
      392 YPKGKDLIVTKEKRYNDPVYITENG--MGDNNAVTEBICXDPORVYFYNQHLSTLKN 449

QY      419 AITODGVVKGYPFAMALIDNLEMSDYGPRFGVTFTDYT-TLKRPKKSALVYK 471
      450 AIAA-GVKVKGYPFAMALIDNLEMSDYGPRFGYIVVDFKGLKRYPGHSLMWRK 502

RESULT 3
US-08-688-988-40
; Sequence 40, Application US/08688988
; Patent No. 6096545
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; APPLICANT: Malboodi, Mohammad A.
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
```

```
FILE REFERENCE: PRL96-03
; CURRENT APPLICATION NUMBER: US/08/688,988B
; CURRENT FILING DATE: 1996-07-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-08-688-988-40

Query Match
Best Local Similarity 36.9%; Score 976; DB 3; Length 509;
Matches 206; Conservative 81; Mismatches 170; Indels 22; Gaps 12;

QY      8 PNDPEWGATTAATVQIEGAVKEGGSGPSIMDTYCHLEPSRTNGANGDVA CDHYHRDEPD 67
      45 PAFVFGTAASAYQVEEMARQGRGFCIMDAFVAIQMILAGNCTADYVDEYHRKEDVG 104
      68 LITKYGAKAYRPSLSMSRIIPLGRLDPVNEBIEFFYSKLDLIRGITTWVLYHMDL 127
      105 IMKMGFDAYRFSISMSRIIPDG--TGKVNQEGVDYNNRLIDYMLQQGITPYANLVHYDL 162
      128 POALHDYRGWLVNEEVQIDFERVYARLCEFERGDRVQNMITINXPMIOAIYATGSNP 187
      163 PLALHQYGLGMLSPKIVGA-FADYAFCFRVPRGDRVKNMFTNEPRVVALGYDNGFAP 221
      188 GRSINKHSTEGNTATEPWLAKQAQIMSHARAVYSRDPRPSQKQIGISLNGDYEPW 247
      222 GRCS--KCPAGDSRTEPYIVTNHLLSHAAVQYREKYQPHQKIRIGILLDFWYEB- 278

QY      248 DSNPRDKEAEERMEFHIGWFANPIFLKKDVPESMKKQIGERLPALTPADPAIINAGET 307
      279 HSDTDDQAAQARDFHIGWFLDPI-TNGRIIPSSMLKIVGNRLPEFS-ADSRMVKGI 336
      308 DRYGMNYYTSQFARHLDG---PVPETD--YLGAIEHOENKDGSPVGEESGLAMLRSCP 361
      337 DYVGINQYTSYWKDGAANNQTPVSYQDDMHVGYVE---RNGVPIGRANSDMLYIVP 392

QY      362 DMFRKHLAVYGLYGP-IYITENGCPCGGEENMCEAVNDPPRIKRFDSHLSISKAI 420
      393 WGNKRAVYTKERYGPNPTILSENGDDPG--NVSIAQDVHVTIRYRYRDIYTELKAI 450

QY      421 TQDGVVKGYPFAMALIDNLEMSDYGPRFGVTFTDYTTLKRPKKSALVKDMPAAR 479
      451 DMGARVAGYFAMSLDNEFMRIGTARGIYVDPNTIKRIPKOSALMFKMLSEKKA 508

RESULT 4
US-09-306-593-9
; Sequence 9, Application US/09306593
; Patent No. 6184018
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, Huizhong
; APPLICANT: Ximenes, Eduardo A.
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
; FILE REFERENCE: 31-9848
; CURRENT APPLICATION NUMBER: US/09/306,593
; CURRENT FILING DATE: 1999-05-06
; EARLIER APPLICATION NUMBER: US 60/084,494
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Costus speciosus
US-09-306-593-9

Query Match
Best Local Similarity 36.5%; Score 964.5; DB 3; Length 476;
```

Best Local Similarity 43.5%; Pred. No. 4,1e-95;  
Matches 204; Conservative 76; Mismatches 168; Indels 19; Gaps 12;

QY 6 ALNDPBMGATTAAYQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDE 64  
DB 9 SFRGRGTFGASAAAYQIEGAMNEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDE 68  
QY 65 DFLRLKYGAKAARFSLMSRIPLGRLDPVNEEGIEFYSKLIDALLRGRITPMYTLVH 124  
DB 69 DVLTLKDLGDSYRFSISMSRIPLKGTLOGGINOEGIQYNDLIDNELLKGRITPMYTLVH 128  
QY 125 WDLPOALHRRYGMVNEEVOJDFERYARLCFERFGDRVOMTTITXPMIOAYGATGS 184  
DB 129 MDVPOALHRRYGMVNEEVOJDFERYARLCFERFGDRVOMTTITXPMIOAYGATGS 187  
QY 185 NAGRSISINSHSTEGNTATPMLAGKQIMSHARAAYVSDRRPSQKQIGISLNDY 244  
DB 188 HAFGRGSIWYGCPRAGSANEPYEVTHNLHLAHANAVYIRDNKYKATONGEIGITLSLWY 247  
QY 245 EPWDSNEPRDKAERMEFHIGFANPIFLKKDYPSMKOLGERLPALTPADPAIINA 304  
DB 248 EPI-SHSEVEYALTRALDMFGWMDPL-VNGDYPIFMALVDRILPFTHAESLIK- 304  
QY 305 GETDFYGMNYYTSQFARHLDPV-----PETDYGAIEHOENKDGSPVEESGLAMLR 359  
DB 305 GSYDFIGINYTSNVAQH--APVTEHPTDPSYFDSVNGSGENGVPIGLQG-SWIFY 361  
QY 360 CPMFRKHLARVYGLYKRP-IYTTENGCPGCEBNMTCEAVNDPFRIRFDSHLSISK 418  
DB 362 YPRLGKELLVYKRRYCNPKIYITENG--TAEVEKEKGVPLHDPKEKVELYTHLAQVLQ 418  
QY 419 AITODGVVGVYFAMALLDLNLEMSDGYGPRFGVTFDY-TTLKRTPKKS 466  
DB 419 AITODGVVGVYFAMALLDLNLEMSDGYGPRFGVTFDY-TTLKRTPKKS 466

## RESULT 5

US-08-688-988-39  
; Sequence 39, Application US/08688988B  
; Patent No. 6096545  
; GENERAL INFORMATION:  
; APPLICANT: Lefebvre, Daniel D.  
; APPLICANT: Malboochi, Mohammad A.  
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS  
; FILE REFERENCE: PRL96-03  
; CURRENT APPLICATION NUMBER: US/08/688,988B  
; CURRENT FILING DATE: 1996-07-31  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 531  
; TYPE: PRT  
; ORGANISM: Prunus avium  
US-08-688-988-39

Query Match 35.0%; Score 924.5; DB 3; Length 531;  
Best Local Similarity 41.1%; Pred. No. 1e-90;  
Matches 197; Conservative 86; Mismatches 182; Indels 15; Gaps 11;

QY 11 FEWGFATAAYQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDEPDL 69  
DB 37 FTGTGTATASQALGAGANIDGRPSIMDAFTNHPEKTIIDSGNDVALIDYHRKEDVALIM 96  
QY 70 TKKGAKAYRSLMSRIPLGRLDPVNEEGIEFYSKLIDALLRGRITPMYTLVHMDLPQ 129  
DB 97 KDMGLAYRFSISMSRIPLGRLDPVNEEGIEFYSKLIDALLRGRITPMYTLVHMDLPQ 156  
QY 130 ALMDRYGMVNEEVOJDFERYARLCFERFGDRVOMTTITXPMIOAYGATGSNAAGR 189  
DB 157 ALDEEYGGVLS-PRIVDFAYAEELCYKERGDRVOMTTITXPMIOAYGATGSNAAGR 215  
QY 190 SS--INHGSTEATATEPMLAKQAQIMSHARAAYVSDRRPSQKQIGISLNDYEPW 247

DB 216 CSSWYDPTCLGDSGTEPYLVTHNLHLAHAAVKLYREKYAQSQEGVICITVWSHFEF- 274

QY 248 DSNBPRDKAERMEFHIGFANPIFLKKDYPSMKOLGERLPALTPADPAIINAGET 307  
DB 275 ASBQKQINASVADALPMFGWMDPL-TRGDDYFQSKRSYKELRPFTEQSSSL-IGSY 332  
QY 308 DFGMNYTSQFARHLDPVNEEVOJDFERYARLCFERFGDRVOMTTITXPMIOAYGATGS 363  
DB 333 DYGVMYARASVADALPMFGWMDPL-TRGDDYFQSKRSYKELRPFTEQSSSL-IGSY 392  
QY 364 FRKHLARVYGLYKRP-IYTTENGCPGCEBNMTCEAVNDPFRIRFDSHLSISKAITQ 422  
DB 393 LYDLVLYTKRKYNDPIMYITTEGMDPEFNNPKISLEQALDMSRIDCYRHLCYLOEAIIT- 451  
QY 423 DGVMYGVYFAMALLDLNLEMSDGYGPRFGVTFDY-TTLKRTPKKSALVLMDFAPAROV 480  
DB 452 EGANVGVYFAMALLDLNLEMSDGYGPRFGVTFDY-TTLKRTPKKSALVLMDFAPAROV 509

## RESULT 6

US-09-306-593-2  
; Sequence 2, Application US/09306593  
; Patent No. 6184018  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; APPLICANT: Ximenes, Eduardo A.  
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from  
; FILE REFERENCE: 31-98us  
; CURRENT APPLICATION NUMBER: US/09/306,593  
; CURRENT FILING DATE: 1999-05-06  
; EARLIER APPLICATION NUMBER: US 60/084,494  
; EARLIER FILING DATE: 1998-05-06  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 657  
; TYPE: PRT  
; ORGANISM: Orpiniomyces sp. PC-2  
US-09-306-593-2

Query Match 34.9%; Score 923.5; DB 3; Length 657;  
Best Local Similarity 36.0%; Pred. No. 1.9e-90;  
Matches 209; Conservative 82; Mismatches 158; Indels 131; Gaps 17;

QY 7 LPNDPBMGATTAAYQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDEPDL 58  
DB 74 LPADFKMGATTAAYQIEGAMNEGGRPSIMDTYCHLEPSR-TNGANGDVACHYHYRDEPDL 133  
QY 59 YHRDEPDLITKYGAKAARFSLMSRIPLG-----GRDPVNEEGIEFYSKLIDALL 112  
DB 134 YHRDEPDLITKYGAKAARFSLMSRIPLG-----GRDPVNEEGIEFYSKLIDALL 192  
QY 113 RRGITPMYTLVHMDLPOALHRRYGMVNEEVOJDFERYARLCFERFGDRVOMTTITXPM 172  
DB 193 KNDIVFVTLVHMDLPOALHRRYGMVNEEVOJDFERYARLCFERFGDRVOMTTITXPM 251  
QY 173 MIOAYGATGS-----SNAP-----GRSINSHSTEGNTATEPMLAKQA 211  
DB 252 MWCVSGYRLGPEKAYRGTGSAFRKLQNSTDLDEGGCSYEIGPYOYSKNSPRLANRV 311  
QY 212 Q-----IMSHARAAYVSDRRPSQKQIGISLNDYEPWMDNEPRDE--- 256  
DB 312 POKLEWVCSHNLILGHAAKAYRGTGSAFRKLQNSTDLDEGGCSYEIGPYOYSKNSPRLANRV 371  
QY 257 ---AABRMFTHIGFANPIFLKKDYPSMKOLGERLPALTPADPAIINAGETDFYGMN 313  
DB 372 NLKYNALAAEFKRGWSDPPMV-GDIPKSYKEMGMDLEFTEBEKKILKSSSDPLGN 430  
QY 314 YTSQFARHLDPVNEEVOJDFERYARLCFERFGDRVOMTTITXPMIOAYGATGSNAAGR 480

Db 431 TTTAHTAAQAQKNEG-----SYIQPTAEANFNDSKKOMDNDCKGRGDWCTIPEPLIG 485  
QY 350 BESSGLAW-----LRSCPDMRKHLARVYGLYKGPITYITENGCPCPG----- 390  
Db 486 SOAGSSMTKRAPTIRVGLNMFSK--RYEGLIKNGIYITERGCQOPNYKVARANDEVTK 542  
QY 391 -----EENMTCEE-----AVNDPRIRYFSDHLSISKAITODGVVKG 429  
Db 543 KYFESIGCPKADYKEDIREDIEDRLBETIMHDTYRIMWOYLKLNLAIAVANDIDVRG 602  
QY 430 YFAALLNLNLEMSDGYGRFGVTFTDY--TTLKRTPKKS 466  
Db 603 YMAWLLDNFEWENGYETRFQWTYIDFYNDKEMKRVKPS 642

RESULT 7  
US-09-122-230-7

Sequence 7, Application US/09122230A  
Patent No. 5973228  
GENERAL INFORMATION:  
APPLICANT: Carlson, et al.  
TITLE OF INVENTION: Coniferin Beta Glucosidase cDNA for Modifying Lignin  
TITLE OF INVENTION: Composition in Plants  
FILE REFERENCE: 50532  
CURRENT APPLICATION NUMBER: US/09/122,230A  
CURRENT FILING DATE: 1998-07-23  
EARLIER APPLICATION NUMBER: U.S. 60/053,566  
EARLIER FILING DATE: 1997-07-24  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 513  
TYPE: PRT  
ORGANISM: Pinus contorta  
US-09-122-230-7

Query Match 34.4%; Score 910; DB 2; Length 513;  
Best Local Similarity 41.6%; Pred. No. 3,6e-89;  
Matches 198; Conservative 83; Mismatches 165; Indels 30; Gaps 13;

QY 8 PNDPEWGATATAAYOIGAVKEGGGSPSIWDYCHLEPSRTMGANGDVAACHYHYDDEDF 67  
Db 31 PSDPMFGTASAYOIEGAVNREDGKQPSITWDLTTHMGRINDSSNGDAVADYHRYMEDIE 90  
QY 68 LITTKYAKAYRSLSMSRIIPLGRLDPVNEEGIEFYSKLDALLRGRTFVWTLVYHMDL 127  
Db 91 LMASLGLDAYRFSISWSRIIPLEG--RGEINMAGIEYVNNLIDALONGIOQFVTLFHFPL 146  
QY 128 PQALHDRVGMVLANVEVQLDPERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGSNAP 187  
Db 149 PKALEDSYGMWLSPOIIN-DEEAYAEICFRAFGRVXKMATVNEPNLFVPLGYGVGFPP 207  
QY 188 GRSS---INKSTSGN-TATEPMLAGKAOIMSHARAVAVYSRDRPSOKOIGISLNGDY 243  
Db 208 TRCAAPHANPLCMQNGCSAEPYLAHVLAAHSAVEKREKQKIQGSGISGLVIAPW 267  
QY 244 YEPWDSNEPRDKEAERMEFHIGWFANPIFLKQDYPSMKKQUGERLPAUTPADFAILN 303  
Db 268 YEPLF-NSPEERSAVDRILSENLRFPLPIVF-GDYQEMERLGRSLPIS-SELSAKL 324  
QY 304 AGEDTDFGMNYYTSQFAHLDGYPVE-TDYL-----GAIHEHENDKGSVGESEG 353  
Db 325 RGSFDTYMGINYYTTLVASTPPLSPDHTQYLYPDSRVYLTGERH-----GVSIGERTG 377  
QY 354 LAWLRSQCPDMEFKHLARVYGLYKRP-IYITENGCPCGEENMTCEAVNDPFRIRYFDSH 412  
Db 378 MDGLFVVPHGIQKIVEYKEFYDNPTIIIANENGPESEESSSTIQENLNDVRRIRFRFGDC 437  
QY 413 LDSISKAITODGVVVKGFAMALLDNLEMSDGYGRFGVTFTDY--TTLKRTPKKSA 467  
Db 438 LSYLSAAL-KNGSDVRGYFVWSLDNFEMAFGYTIRGLVHVDPIQKQKPKLSA 492

RESULT 8  
US-09-431-470-4

Sequence 4, Application US/09431470  
Patent No. 6433249  
GENERAL INFORMATION:  
APPLICANT: Duvick, Jon  
APPLICANT: Simmons, Carl R.  
APPLICANT: Craets, Oswald R.  
APPLICANT: Folkerts, Otto  
TITLE OF INVENTION: The Use of Beta-Glucosidase to Enhance  
Disease Resistance to Insects in Crop Plants.  
FILE REFERENCE: 5718-43  
CURRENT APPLICATION NUMBER: US/09/431,470  
CURRENT FILING DATE: 1999-11-01  
PRIOR APPLICATION NUMBER: 60/107,920  
PRIOR FILING DATE: 1998-11-10  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 563  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-431-470-4

Query Match 34.1%; Score 902; DB 4; Length 563;  
Best Local Similarity 39.5%; Pred. No. 3,1e-88;  
Matches 195; Conservative 98; Mismatches 169; Indels 32; Gaps 15;

QY 8 PNDPEWGATATAAYOIGAVKEGGGSPSIWDYCHLEPSR-TMGANGDVAACHYHYDEDF 66  
Db 76 PSDPMFGTASAYOIGAVNEDGKGSNMDHCFHNPERRIMDSNADIGANSYMYKTIV 135  
QY 67 LITTKYAKAYRSLSMSRIIPLGRLDPVNEEGIEFYSKLDALLRGRTFVWTLVYHMD 126  
Db 136 RLTKEMGMDAYRFSISMPRLPKGYVEGGINDGIDYKRLNLNLENGIBPVTLFHM 195  
QY 127 LPQALHDRVGMVLANVEVQLDPERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGS 184  
Db 196 VPQALBEKKGGLFDKTKQKILVNDYKNAFKYCFDNGDKXNMLTFNEPQTFPSFSYGTGV 255  
QY 185 NAGRSS--INKSTSGNATEPMLAGKAOIMSHARAVAVYSRDRPSOKOIGISLNGD 242  
Db 256 FARGRSPGIDCAIPTGNSLVEPYIAGHNILAAHAAVDLYNKYYK-GENGRIGLAFDW 314  
QY 243 YEPWDSNEPRDKEAERMEFHIGWFANPIFLKQDYPSMKKQUGERLPAUTPADFAIL 302  
Db 315 GRYPYGTSP-LDEQAKERSMDINLGFLEPV-VRGYPFSMSLSARELPFSDDKQEK 372  
QY 303 NAGETDFYGMNYYTSQFAHLD-----GVPEFTDYLGAIHEHENDKGSVGESEGLAWL 357  
Db 373 -VGSYNNLGINYYTISFSKHIDISPKYSPVLNTDDAYASQF-TYGPDKRIGPMPGNPMI 430  
QY 358 RSCPDMEFKHLARVYGLYKRP-IYITENGCPCPG-----EENMTCEAVNDPFRIRYFDS 411  
Db 431 YIYBEGLKQILIMTKKRYNPNPYITENGI---GDVDTKEKELPMEEAALNDYKRLDYTOR 487  
QY 412 HDSISKAITODGVVVKGFAMALLDNLEMSDGYGRFGVTFTDY--TTLKRTPKKSA 467  
Db 488 HISTLKESIDL-CANHGAFMSLNDNPFEMVAGYIERGYIYVDRKNMT---RYMKESA 543  
QY 468 LVTKDMFAARQRYK 481  
Db 544 KWLKEPNTAKKPSK 557

## RESULT 9

US-08-688-988-38  
Sequence 38, Application US/08688988B  
Patent No. 6096545  
GENERAL INFORMATION:  
APPLICANT: Lefebvre, Daniel D.  
APPLICANT: Malboochi, Mohammad A.  
TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS

FILE REFERENCE: PPL96-03  
 CURRENT APPLICATION NUMBER: US/08/688,988B  
 CURRENT FILING DATE: 1996-07-31  
 NUMBER OF SEQ ID NOS: 48  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 38  
 LENGTH: 531  
 TYPE: PRT  
 ORGANISM: Manihot esculenta  
 US-08-688-988-38

Query Match 33.9%; Score 896; DB 3; Length 531;  
 Best Local Similarity 40.3%; Pred. No. 1,2e-87;  
 Matches 190; Conservative 90; Mismatches 172; Indels 20; Gaps 11;

QY 8 PNDFFGFAATAVOIEGAVKEGGRPSIMPTYCHLEPSR-TNGANGVACDHYHREDEP 66  
 DB 35 PDDFIFGTAATAOIEGATAGKAPSWDIFSKETPDRLDSDGVAADFYNYIQDI 94  
 QY 67 DLTLYGAKAYRSLWSRIIPGGRDLPVNEGIEFYSKIDALLRGITTPWVTLVHM 126  
 DB 95 KAVKMGFNAFRMSISRSVIPSGRRREGVNEBGIOFYNDVINELISNGLEPVTIIFHM 154  
 QY 127 LPOALHRYGGMVNEBVDLFEFVYARLCEPFGDRVQNMITTXMPWIAIYGATGSA 186  
 DB 155 TPQALODKYGGFLS-RDIYVDYLQYADLLFERFGDRVKPMWTFNEPSAYVGFAMDGVFA 213  
 QY 187 RGRSS--INKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 244  
 DB 214 PGCSWVNNQCLAGSATEPYIVAHNLISHAAVAQYKRYOGTKIGITLFTFWY 273  
 QY 245 EPDUSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPALTTPADFAILNA 304  
 DB 274 EPL-SOSKVDVQAKTALDPMFGLMMDPMYGR-YRPTWDLADGKLGITDESQULR- 330  
 QY 305 GETDFYGMNYYISQFAHLDGVPETD----YL-GAIEHOENKDGSPVGEESGLAWL 357  
 DB 331 GSYDFVQLQYTYAVAE---PIPVDPKPRRYKTDGCVNATPYDLNGLIGQAVSSWF 386  
 QY 358 RSGEDMPRKLARVYGLYCKP-IYITENGCPGCEENMTCEAVNDPFRIRYDSHLSI 416  
 DB 387 YIFPKGIRHFLNTKOTYNDPVLYVTEGNDVNNESQPIEALQDDPFIYSYKKNHMA 446  
 QY 417 SKAITODGVVVKGFAMALDNLWSDGYGPRFGVTFTDY-TLLKRTPKSA 467  
 DB 447 LGSLLKMYGVKLCGYFAMSYLDNFEKNIIGYRSGLYVVDYKNNLTRPKSA 498

RESULT 10  
 US-08-914-375C-72  
 Sequence 72, Application US/08914375C  
 Patent No. 6377893  
 GENERAL INFORMATION:

APPLICANT: Steven A. Benner  
 Applications of Protein Structure Predictions  
 NUMBER OF SEQUENCES: 74  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Steven A. Benner  
 STREET: 1501 NW 68th Terrace  
 CITY: Gainesville  
 STATE: FL  
 COUNTRY: United States  
 ZIP: 32605-4147  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch diskette  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Macintosh 7.0  
 SOFTWARE: Microsoft Word  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/914,375C  
 FILING DATE: 19-Aug-1997  
 CLASSIFICATION: 702/20  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 352 392 7773  
 TELEFAX: 352 331 0462  
 INFORMATION FOR SEQ ID NO: 72:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 459  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: amino acid  
 ORIGINAL SOURCE:  
 ORGANISM: Trifolium repens  
 FEATURE:  
 OTHER INFORMATION: bgl's trip non-cyanogenic strand-glucosidase precursor  
 SEQUENCE DESCRIPTION: SEQ ID NO: 72:  
 US-08-914-375C-72

Query Match 33.3%; Score 880.5; DB 4; Length 459;  
 Best Local Similarity 40.3%; Pred. No. 4.6e-86;  
 Matches 186; Conservative 80; Mismatches 177; Indels 19; Gaps 8;

QY 6 ALPNDGFWFATAVAOIEGAVKEGGRPSIMPTYCHLEPSR-TNGANGVACDHYHRYDE 64  
 DB 5 SFRGFIIFGAGSAYVPEGAVNEGGRPSIMPTFTKYEKIRDSNADITVDQHYRYE 64  
 QY 65 DFDLTKYAKAYRSLWSRIIPGGRDLPVNEGIEFYSKIDALLRGITTPWVTLVH 124  
 DB 65 DVGIMKQNMDSYRFSISMPRIIPKGLSGINHEGIKYNNMLINELLANGIQPFVTLFH 124  
 QY 125 WDLPOLHRYGGMVNEBVDLFEFVYARLCEPFGDRVQNMITTXMPWIAIYGATGS 184  
 DB 125 WDLPOLHRYGGMVNEBVDLFEFVYARLCEPFGDRVQNMITTXMPWIAIYGATGS 183  
 QY 185 NAFGRSSINKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 244  
 DB 184 NAFGRSSINKHSTEGTATEPMLAGKAQIMSHARAVAVSRDPSPQKQIGISLNDY 243  
 QY 245 EPDUSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPALTTPADFAILNA 304  
 DB 244 MPDLDNISIDIKAAERSLDFOFGLFMEQL-TTGDSKSNRRIVKXNLPYFSKRESSLVN- 301  
 QY 305 GETDFYGMNYYISQFAHLDG----PVPETDYLGAIEHOENKDGSPVGEESGLAWL 359  
 DB 302 GSYDFVQLQYTYAVAE---PIPVDPKPRRYKTDGCVNATPYDLNGLIGQAVSSWF 358  
 QY 360 CPDMFPRKLARVYGLYCKP-IYITENGCPGCEENMTCEAVNDPFRIRYDSHLSI 413  
 DB 359 YIFPKGIRHFLNTKOTYNDPVLYVTEGNDVNNESQPIEALQDDPFIYSYKKNHMA 418  
 QY 414 DSISKAITODGVVVKGFAMALDNLWSDGYGPRFGVTFTD 455  
 DB 419 YYIRSAI-RAGSNVKGFYAMSLDCNEMWAFGTVRFGLNFDV 459

RESULT 11  
 US-09-431-470-2  
 Sequence 2, Application US/09431470  
 Patent No. 6433249  
 GENERAL INFORMATION:

APPLICANT: Duvick, Jon  
 APPLICANT: Simmons, Carl R.  
 APPLICANT: Craata, Oswald R.  
 APPLICANT: Folkerts, Otto  
 TITLE OF INVENTION: The Use of Beta-Glucosidase to Enhance  
 TITLE OF INVENTION: Disease Resistance to Insects in Crop Plants.  
 FILE REFERENCE: 5718-43  
 CURRENT APPLICATION NUMBER: US/09/431,470  
 CURRENT FILING DATE: 1999-11-01  
 PRIOR APPLICATION NUMBER: 60/107,920  
 PRIOR FILING DATE: 1998-11-10  
 NUMBER OF SEQ ID NOS: 4  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 566  
 TYPE: PRT

ORGANISM: Zea mays  
US-09-431-470-2

Query Match 33.2%; Score 879; DB 4; Length 566;  
Best Local Similarity 38.3%; Pred. No. 9,4e-86;  
Matches 192; Conservative 94; Mismatches 183; Indels 32; Gaps 14;

1 MPESIALPNDFWGFATPAVOIGCAVKEGGRPSIMWTYCHLEPSR--TNGANDVACDXY 59  
2 IPRDWFPSDFTFGAATSAVOIGCANWEDGSGESNMFCHNPERLLDSNDIDANXY 131  
3 HRYDEDPLLTXYGAKVRFSLWSRIIPLGRIIDPVNEBIEFYSKLDALIRGITPW 119  
4 HMYKTDLVLEKMGMDAYRFSISMPRIIPKGTKEGGINPDGIKYRMLIMILENGIEPY 191  
5 VLYHMDLPOLHRYGWLNVVEVQL--DPERYARLCFERFGDGVONWITINXPMIOAI 177  
6 VTIFFMDVPOALEEKYGGFLDKSHKSIYEDYTYPAKCEDFGDKVKNWLTNEPQTFIS 251  
7 YGIATGSAAPGRSS--INKHSTEGNTATEPWLAKQAQIMSHARAVAVYSRDPRSQGQI 235  
8 FSYGTGVFAPGRCSGLDCAIPTGSLVEPYTAGHINILAAEAVDLYNKHYKDDT-RI 310  
9 GISTNGDYEPWDSNEPRDKEAERMEFHIGWPNPIFLKQDYPESKKQIGRLPALT 295  
10 GLAFDVGWVPGTGF-LDKQAEERSWDINLGWLEPV-VRGDYPFSMRSLARELPEFK 368  
11 PADPAILNGETDFPGMNYTTSOPARHLD-----GPVETDYLGAIHENKQSGPYGE 350  
12 DEQEKEL-AGSYNMGILNTYTSRFSKNIDISPNYSPLNTDVAVSQE-VNPGDQKPIPG 426  
13 ESGLAWLRSCPDWFRKHLARVYGLYGR-PITYTENG-----CPGGEENMTCEAVN 401  
14 PKMNPPIIYMPBELKXOLIMMKKYGNPPIYITENGIGDVTKEPLP-----MEALIN 480  
15 DPFRIIRYSHLDSIKAITQGVVYKGFAMALLDNLEMSGYGRFGVTFD-YTLX 460  
16 DYKRLDIOHRIYTLKESIDL-GSNVQGFANSLDNFPMFAGFTERYGIYVVDNNNCT 539  
17 RTPKKSALVLDKMFARQRYK 481  
18 RYKESAKMLKEPNTAKKPSK 560

RESULT 12  
US-08-688-988-34  
Sequence 34, Application US/08688988B  
Patent No. 6096545  
GENERAL INFORMATION:  
APPLICANT: Lefebvre, Daniel D.  
APPLICANT: Malloobi, Mohammad A.  
TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS  
FILE REFERENCE: PPL96-03  
CURRENT APPLICATION NUMBER: US/08/688, 988B  
CURRENT FILING DATE: 1996-07-31  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 34  
LENGTH: 524  
TYPE: PRT  
ORGANISM: Arabidopsis Thaliana  
US-08-688-988-34

Query Match 32.7%; Score 863.5; DB 3; Length 524;  
Best Local Similarity 39.8%; Pred. No. 3,9e-84;  
Matches 196; Conservative 73; Mismatches 198; Indels 25; Gaps 13;

6 ALPNDFWGFATPAVOIGCAVKEGGRPSIMWTYCHLEPSR--TNGANDVACDXY 65  
39 SPEEGFLGTATDAYVEGAINETCRGPMALMDICRRYPERCNDNDGVAVDFHRYKD 98  
66 FDLITRYGAKVRFSLWSRIIPLGRIIDPVNEBIEFYSKLDALIRGITPWITLXW 125

99 IOLMKNLNTDAPRMSIAMPRIIPPHGRKEKGVSAQVGFYHDLIDBLINKNITPFWTFHW 158  
106 DLPOALHRYGWLNVVEVOLDPERYARLCFERFGDGVONWITINXPMIOAIYVATGSN 185  
109 DTPDLDEYVGGLS-ERIVKDFREYADVFQRYGGKVKHMITNEPFWFLHAGYVGKK 217  
116 AGRSS--INKHSTEGNTATEPWLAKQAQIMSHARAVAVYSRDPRSQGQIGISLNGY 243  
118 AGRCSYVNAKCODGSGYEAVLYTHNLLISHAEVAV-RCKECKGKIGIAHSPM 276  
119 YEPWDSNEPRDKEAERMEFHIGWPNPIFLKQDYPESKKQIGRLPALTPADPAIIN 303  
120 FEADHLDOSOGASIDALDFILGMLDITTF-GDYQIKMDIYGHRLPFTTEQOKAKK 335  
121 AGEDYFGMNYTTSOPARHLDGPVE-----TDYLAGIHENKQSGPYGESGLAWLR 358  
122 A-STDFVLNNTYTSVFSNHLKPPDSKPRMWOBSL-ITWESKNAQVYAGISPLTALN 392  
123 SCPDWFRKHLARVYGLYGR-PITYTENGCPGGEENMTCEAV-----NDPFRIYFDSHL 413  
124 VYSRGFRSLKXIKDKYANPEIMENGY--GEE-LGASDSVAVGTADHNRKTYLQRL 448  
125 DISKAITQGVVYKGFAMALLDNLEMSGYGRFGVTFDY-TTLKRTPKKSALVLD 472  
126 LMOEAVCIDKVNVTGYFVWSLDFWQDGYKRRFGLYYVDPFONILTRYEKESGKYKD 508  
127 MFARQRYKVA 484  
128 FLS--QVAPSA 518

RESULT 13  
US-09-914-841A-1  
Sequence 1, Application US/09914841A  
Patent No. 6645750  
GENERAL INFORMATION:  
APPLICANT: Amano Enzyme Inc.  
TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE  
FILE REFERENCE: 066072  
CURRENT APPLICATION NUMBER: US/09/914, 841A  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: P. Hel. 11-056299  
PRIOR FILING DATE: 1999-03-04  
PRIOR APPLICATION NUMBER: PCT/JP00/01242  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 429  
TYPE: PRT  
ORGANISM: Camelilia var. sinensis  
US-09-914-841A-1

Query Match 32.6%; Score 861; DB 4; Length 429;  
Best Local Similarity 42.6%; Pred. No. 5,2e-84;  
Matches 184; Conservative 73; Mismatches 159; Indels 16; Gaps 11;

47 TNGANDVACDHYRYDEDPLLTXYGAKVRFSLWSRIIPLGRIIDPVNEBIEFYSK 106  
2 SNGSTGVADDFYHRYKEDVYLVKFIGLDFRMSISWARYLPRGKLSGVNKEGIAFYNN 61  
107 LIDALIRGITPWITLXWDLPOALHRYGWLNVVEVOLDPERYARLCFERFGDGVONW 166  
62 VINDLISKIGIPFITIFHMDLPOLAEDEYGGFLSPHYN-DPRPAELCFERFGDGRKHW 120  
167 ITINXPMIOAIYVATGSNAPGR-SSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSR 225  
121 ITWNEPWSYSGGYDAGLAPERCAPFAFCPKNGSGTEPIYVTHNLLISHAAVVKYKE 180  
226 DFRPSQGQIGISLNGDYEPWDSNEPRDKEAERMEFHIGWPNPIFLKQDYPESKKK 285  
181 KYAYQKQIGITIVTYWMIY-SNSKADKDAACALDFMYGWFLEPLSF-GEYPKSMR 238

```

Oy 286 OLGEELPALTPADPHILINGETDFEGMNYTYSQFAHLHOGPVEETOYLGAIHQGN--- 342
Db 229 LVGKRLPRT-KEQMLVKGSPDFGLNYYIANV--LVNPTNSNVLSITDLSNQRA 255
Oy 343 -KDGSVPVEESGLAWLRSCPDWFRKHLARVYGLYGRP-IYITENGCPCEENMTCEAV 400
Db 296 FRNGVALICRPVCVPAFMYPKGLKDLVYTKERKYNDPVIYITENG--MGDNNVTTTEBGI 353
Oy 401 NDPFRIRFDSHLDSISKAITDDGVVYKGYFAMALLDNLEMSDGYCPRRGVFTDYTT-L 459
Db 354 KDPRVVYVYNNHLLSLKNAIAA-GVKVKGYFTWAFIDNEFWLSGYQRFGIYVVDKPDGL 412
Oy 460 KRTPKSAALVLK 471
Db 413 KRIYKHSALMEK 424

```

RESULT 14  
US-08-914-375C-69  
; Sequence 69, Application US/08914375C  
; Patent No. 6377893  
; GENERAL INFORMATION:

```

1      NUMBER OF SEQUENCES: 74
2      APPLICATIONS OF Protein Structure Predictions
3      CORRESPONDENCE ADDRESS:
4      ADDRESSEE: Steven A. Benner
5      STREET: 1501 NW 68th Terrace
6      CITY: Gainesville
7      STATE: FL
8      COUNTRY: United States
9      ZIP: 32605-4147
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE: 3.5 inch diskette
12     COMPUTER: Apple Macintosh
13     OPERATING SYSTEM: Macintosh 7.0
14     SOFTWARE: Microsoft Word
15     CURRENT APPLICATION DATA:
16     APPLICATION NUMBER: US/08/914,375C
17     FILING DATE: 19-Aug-1997
18     CLASSIFICATION: 702/20
19     TELECOMMUNICATION INFORMATION:
20     TELEPHONE: 352 392 7773
21     TELEFAX: 352 331 0462
22     INFORMATION FOR SEQ ID NO: 69:
23     SEQUENCE CHARACTERISTICS:
24     LENGTH: 476
25     TYPE: amino acid
26     TOPOLOGY: linear
27     MOLECULE TYPE: amino acid
28     ORIGINAL SOURCE:
29     ORGANISM: Brassica napus
30     FEATURE:
31     OTHER INFORMATION: myro brana myrosinase precursor (E.C. 3.2.3.1
32     SEQUENCE DESCRIPTION: SEQ ID NO: 69:
33     US-08-914-375C-69

```

	Query Match	32.5%	Score 858.5	DB 4:	Length 476:
	Best Local Similarity	40.1%	Pred. No. 1.2e-83:		
	Matches 194;	Conservative	87;	Mismatches 156;	Indels 47; Gaps 16
QY	10	D F E W G F A T A A Y Q I E G A V K R G S P I M P T Y C H L E P S R N T G - - A N G D V A C D H Y R Y D E D F D	67		
	:	:::::	::	: ::	:: ::
Db	9	D F I G V A S S A V Q I E G G - - - R G R G V V M V G F S H R Y E K A G S D L K N G D T T C E S T T R K Q V Y D	65		
QY	68	L L T Y T G A K A Y F S L S W S R I I L P L G R L D P V N E E G I E F Y S K L I D A L I R G I T P W T L Y H M D L	127		
	:::				
Db	66	V M G E I N A T G Y F F S F A M S R I I P K G K X S R G V N O G S D L Y H K L I D A L L E K N I T P E V T L F P H M D L	125		
QY	128	P Q A L H D R G R G L N Y E E V O L D F E R V Y A R L C F E R R G D B V Q M I T T N X P M I A I G Y A T G S N A P	187		
	:	::: ::			
Db	126	P Q T I D E Y E G G L D R Q I O - D F K D V A D L C F K E G K V K M I T T N Q L Y T P T R I A L G T D A P	184		

[illegible]

RESULT 15  
ITS-08-68B

; Sequence 41, Application US/08688988H  
; Patent No. 6096545

```

1  APPLICANT: Lefebvre, Daniel D.
2  APPLICANT: Malboochi, Mohammad A.
3  TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
4  FILE REFERENCE: P196-03
5  CURRENT APPLICATION NUMBER: US/08/688,988B
6  CURRENT FILING DATE: 1996-07-31
7  NUMBER OF SEQ ID NOS: 48
8  SOFTWARE: FastSeq for Windows Version 3.0
9  SEQ ID NO 41
10 LENGTH: 548
11 TYPE: PR1
12 ORGANISM: Brassica Napus
13 US-08-688-988-41

```

Query Match	32.5%	Score 858.5	DB: 3	Length 548
Best Local Similarity	40.1%	Pred. No. 1,5e-83		
Matches 194	Conservative 87	Mismatches 156	Indels 47	Gaps 16
Oy	10	DEEMGPTAAVOIGVAKVEGRGSPIMDTYHLEPSRNG--ANGDVAChHYRDEDFD	67	
Db	48	DFIVGVASSAAQIIGG--RGRGVNVWDGFSHRYPEKASDCLKGDTTCESYTRMOKVD	104	
Oy	68	LITTKGAKAYEFSWSRRIIPLGRLDPVNEEGIEFYSKLTIDALLRGITPMVTLTYHMDL	127	
Db	105	VMGELNATGVFSSTAWSRRIIPKGVKSRGVNGGIDYHKLIDALLEKQITPVTLLFHMDD	164	
Oy	128	POALHDRVGYMLNTEVEVOLDEPERYALCFEEFGDGRVQWMTINXPMWIAIYGATGSNAP	187	
Db	165	PQTADDEYEGFLDQIITD-DPKDYADLCFEKFGKVKWMTIINOLYTPPTGYAIGTAP	223	
Oy	188	GRSS---INKHST-EGNTATEPMWJAKQAQIIMSHARAVALVYSRDERSPOQSIGISLNDGY	243	
Db	224	GRCPMWDTRKRCYCGGNSSTPEPIYVAINNQLAHATVVDLYTTKK-FQKGIAPMIRW	282	
Oy	244	YEPWDSNEPRDKEAERMEFHIGMFANPIFLKXDYESMKKOLGERLPALTTPADFAILN	303	
Db	283	FLPPEDSPPASIEAAERNNQFFHOMYEP-L-TKRYRPIMKQIYGSRLPNPTTEE-ALVY	340	
Oy	304	AGETDPYGMNYYTSQFAHLDGVPV-ET-----DYLGAIHHEQENKD	344	
Db	341	AGSYDPLGLNLYVYVQAQPKENPYRSEFHTMTAMDAGVCLTYDNRSGEFLGPLV-EDKVN	399	
Oy	345	GSPLVGEBSGLAMLRSCPDMPFKHLLARVYGLYGRK-IYITENGCCPCGGEANNTCCEAVNDR	403	

Db	400	GNQYYPKGIYYM--DYFKTK-----YGDPLIYVTENGFSFPSSENR--EQAIADY	447
QY	404	FRIRYPSHSDSISKAITODGVVKGYPFAWALLDNLEMSDYGPRFGVFTDYTL-KRT	462
Db	448	KRIDYLCSHLCFLRKVIKEKGVNVRGTFPAWALGDNIEFCCKGFTVRFGLSYVNWEDLDDRN	507
QY	463	PKKS	466
Db	508	LKES	511

Search completed: March 25, 2004, 15:57:43  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 25, 2004, 15:56:42 ; Search time 46 Seconds  
(without alignments)  
2753.137 Million cell updates/sec

Title: US-10-026-140-2

Sequence: 1 MPESLALPNDPFWGFATAY.....KSAVLKDMFAARQYKVA 484

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1065169 seqs, 26161801 residues

Total number of hits satisfying chosen parameters: 1065169

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubppa/PCF\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubppa/PCFUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2642	99.9	484	US-10-026-140-2	Sequence 2, Appl1
2	1210	45.8	490	US-10-369-493-13405	Sequence 13405, A
3	1058.5	40.0	491	US-10-424-599-169310	Sequence 169310, A
4	1058.5	40.0	511	US-10-425-114-43168	Sequence 43168, A
5	999.5	37.8	515	US-10-425-114-55300	Sequence 55300, A
6	999.5	37.8	516	US-10-424-599-243532	Sequence 243532, A
7	999.5	37.8	519	US-10-425-114-48403	Sequence 48403, A
8	999.5	37.8	519	US-10-425-114-55444	Sequence 55444, A
9	999	37.8	527	US-10-425-114-56976	Sequence 56976, A
10	999	37.8	527	US-10-425-114-70165	Sequence 70165, A
11	999	37.8	527	US-10-425-114-71098	Sequence 71098, A
12	997	37.7	523	US-10-424-599-213409	Sequence 213409, A
13	997	37.7	552	US-10-425-114-55483	Sequence 55483, A
14	996.5	37.7	503	US-10-424-599-147332	Sequence 147332, A
15	986	37.3	510	US-10-424-599-219958	Sequence 219958, A

16	986	37.3	522	US-10-425-114-48576	Sequence 48576, A
17	977	37.0	531	US-10-424-599-214089	Sequence 214089, A
18	972	36.8	525	US-10-425-114-50107	Sequence 50107, A
19	963	36.4	521	US-10-425-114-59902	Sequence 59902, A
20	961.5	36.4	515	US-10-424-599-175202	Sequence 175202, A
21	960	36.3	528	US-10-425-114-65044	Sequence 65044, A
22	959	36.3	531	US-10-425-114-40755	Sequence 40755, A
23	956.5	36.2	526	US-10-424-599-213328	Sequence 213328, A
24	953	36.0	524	US-10-424-599-281853	Sequence 281853, A
25	943.5	35.7	545	US-10-195-781A-4	Sequence 4, Appl1
26	941.5	35.6	506	US-10-424-599-181025	Sequence 181025, A
27	934	35.3	584	US-10-425-114-64343	Sequence 64343, A
28	930.5	35.2	574	US-10-425-114-44487	Sequence 44487, A
29	928.5	35.1	527	US-10-424-599-283766	Sequence 283766, A
30	925	35.0	523	US-10-425-114-54197	Sequence 54197, A
31	922	34.9	541	US-10-425-114-50689	Sequence 50689, A
32	914	34.6	510	US-10-425-114-45687	Sequence 45687, A
33	914	34.6	512	US-10-425-114-44614	Sequence 44614, A
34	914	34.6	513	US-10-424-599-263085	Sequence 263085, A
35	914	34.6	513	US-10-425-114-45855	Sequence 45855, A
36	912.5	34.5	565	US-10-425-114-38182	Sequence 38182, A
37	902	34.1	478	US-10-425-114-54717	Sequence 54717, A
38	884	33.4	508	US-10-425-114-46495	Sequence 46495, A
39	884	33.4	567	US-10-425-114-65966	Sequence 65966, A
40	884	33.4	571	US-10-425-114-72958	Sequence 72958, A
41	884	33.4	571	US-10-425-114-73036	Sequence 73036, A
42	884	33.4	572	US-10-425-114-64277	Sequence 64277, A
43	884	33.4	572	US-10-425-114-72696	Sequence 72696, A
44	882	33.4	571	US-10-425-114-66560	Sequence 66560, A
45	879	33.2	567	US-10-425-114-65837	Sequence 65837, A

## ALIGNMENTS.

RESULT 1  
US-10-026-140-2  
; Sequence 2, Application US/10026140  
; Publication No. US20030114330A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC697  
; CURRENT APPLICATION NUMBER: US/10/026,140  
; CURRENT FILING DATE: 2002-03-26  
; NUMBER OF SEQ. ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 484  
; TYPE: PRT  
; ORGANISM: Trichoderma reesei  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(484)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
; US-10-026-140-2

Query Match 99.9%; Score 2642; DB 14; Length 484;

Best Local Similarity 100.0%; Pred. No. 9, 4e-261; Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPESLALPNDPFWGFATAYQIEGAVGEGRGPSIMDYTCHELPSPRTNGANGVACDHYH	60
DB	1	MPESLALPNDPFWGFATAYQIEGAVGEGRGPSIMDYTCHELPSPRTNGANGVACDHYH	60
QY	61	RYEDDFLLTKYAKAYRFLSLMSRIIPLCGRDPPVNEBIEFFYSKIDLLRGTTPWV	120
DB	61	RYEDDFLLTKYAKAYRFLSLMSRIIPLCGRDPPVNEBIEFFYSKIDLLRGTTPWV	120

QY 121 TLVHMDLFOALHDIRYGMVNLVEVQLDPERYARLCFERFGDRVQNMITINXPMIOAIYGY 180  
Db 121 TLVHMDLFOALHDIRYGMVNLVEVQLDPERYARLCFERFGDRVQNMITINXPMIOAIYGY 180  
QY 181 ATGSAAPRRSSINKESTGNTATTEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLN 240  
Db 181 ATGSAAPRRSSINKESTGNTATTEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLN 240  
QY 241 GDYEPMSNPRDKEAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFA 300  
Db 241 GDYEPMSNPRDKEAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFA 300  
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQENKDGSPVGEESGLAMLRSC 360  
Db 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQENKDGSPVGEESGLAMLRSC 360  
QY 361 PDMPKHLARVYGLYGRITTYTENCPCPGSEENMTCEAVNDPFRIRFDSHLSISKAI 420  
Db 361 PDMPKHLARVYGLYGRITTYTENCPCPGSEENMTCEAVNDPFRIRFDSHLSISKAI 420  
QY 421 TODGVVVGYPFAMALLDNLEMSDGYGPRFGVTFTYTLKRTPKKSALVLDMPAARQV 480  
Db 421 TODGVVVGYPFAMALLDNLEMSDGYGPRFGVTFTYTLKRTPKKSALVLDMPAARQV 480  
QY 481 KVAA 484  
Db 481 KVAA 484

## RESULT 2

US-10-369-493-13405  
; Sequence 13405, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 13405  
; LENGTH: 490  
; TYPE: PRT  
; ORGANISM: Aspergillus nidulans  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(490)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-13405

Query Match 45.8%; Score 1210; DB 15; Length 490;  
Best Local Similarity 47.3%; Pred. No. 2,1e-114;  
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;

QY 7 LPNDFEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 66  
Db 7 LPNDFEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 66  
QY 67 DLTKYGAKAARFSLMSRIIPLGRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHMD 126  
Db 67 ALMSYGNNAARFSLMSRIIPLGRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHMD 126  
QY 127 LPQALHDIRYGMVNLVEVQLDPERYARLCFERFGDRVQNMITINXPMIOAIYGYATGSA 186  
Db 127 LPQALHDIRYGMVNLVEVQLDPERYARLCFERFGDRVQNMITINXPMIOAIYGYATGSA 186

QY 187 PGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLNQDYEP 246  
Db 187 PGRSSINKHSTEGNTATEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLNQDYEP 246  
QY 247 WDSNEPPDKDAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFAIINAGE 306  
Db 247 WDSNEPPDKDAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFAIINAGE 306  
QY 307 TDFPGMNYVTSQFARHLDPVPETDYLGAIHEHQENKDGSPVGEESGLAMLRSCQPMFRK 366  
Db 307 TDFPGMNYVTSQFARHLDPVPETDYLGAIHEHQENKDGSPVGEESGLAMLRSCQPMFRK 366  
QY 367 HLAARVYGLYGRITTYTENCPCPGSEENMTCEAVNDPFRIRFDSHLSISKAI 426  
Db 367 HLAARVYGLYGRITTYTENCPCPGSEENMTCEAVNDPFRIRFDSHLSISKAI 426  
QY 427 VKGYFAMALLDN-----LEMSDGYGPRFGVTFTYTLKRT--PKKSAL 468  
Db 427 VKGYFAMALLDN-----LEMSDGYGPRFGVTFTYTLKRT--PKKSAL 468  
QY 469 VLKMF 474  
Db 469 VLKMF 474

## RESULT 3

US-10-424-599-169310  
; Sequence 169310, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihui  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; PRIOR FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 169310  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_123901C.1.pcp  
US-10-424-599-169310

Query Match 40.0%; Score 1058.5; DB 12; Length 491;  
Best Local Similarity 46.7%; Pred. No. 6.5e-99;  
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

QY 8 PNDPEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 67  
Db 8 PNDPEWGFATAAYQIEGAVGEGRGPISIMDTYCHLEPSRTNGANGVACDHRYRDEDF 67  
QY 68 LITRYGAKAARFSLMSRIIP--LGGRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHMD 125  
Db 68 LITRYGAKAARFSLMSRIIP--LGGRLDPVNEEGIEFYSKIDALLRRGITPMVTLYHMD 125  
QY 142 DLPLHLEHSGWGLNKQILEY--FAVYADTFEAFSGDRVQNMITINXPMIOAIYGYATGSA 200  
Db 142 DLPLHLEHSGWGLNKQILEY--FAVYADTFEAFSGDRVQNMITINXPMIOAIYGYATGSA 200  
QY 201 APGRSSINKESTGNTATEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLNQDYEP 245  
Db 201 APGRSSINKESTGNTATEPMLAGKQIMSHAAVAVYSDPFPQKQIGISLNQDYEP 245  
QY 246 PMSNPRDKEAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFAI--LNA 304  
Db 246 PMSNPRDKEAERMEFHIGWFANPIFLKKDYPPSMKKQLGERLPALTPADFAI--LNA 304  
QY 305 GETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQ--ENKDGSPVGEESGLAMLRSCP 361  
Db 305 GETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHQ--ENKDGSPVGEESGLAMLRSCP 361

```

Db      311 --LDIFGLNHTYSRFISHVTECAEENHY--KVQEMERIVVEGQALIGEKASMLYVVP 367
Qy      362 DMFRKHLARYGGLYGRKPIYITENGCCPCGEBENMTCEAVNDPFRIRYFDSHLSISKAIT 421
Db      368 WGLRKILINYSOKYATPIFTENGMDDEDNDNLPLHMLDKLRYRFFKGLASVAQAI- 426
Qy      422 QDGVVVGYPFAMALLDNLWMSDGYGPRGVTFTDYTT-LKRTPKKSA 467
Db      427 KDGADVGVGFAMSLDNLFEWAQGYTKRFGLVYVDYKNGLSRHPKSSA 473

```

# RESULT 4

```

US-10-425-114-43168
; Sequence 43168, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingtong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabacka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43168
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700748475_FLI.pep
US-10-425-114-43168

```

Query Match 40.0%; Score 1058.5; DB 12; Length 511;

Best Local Similarity 46.7%; Pred. No. 6.9e-99; Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

Qy      8 PNDPEWGFATAAYOIEGAVEGGRGPSIMDTYCHLEPSRTNGANGVACDHRYHRYDEPD 67
Db      45 PNFIFGVATSAAYOIEGAVEGGRGPSIMDTYCHLEPSRTNGANGVACDHRYHRYDEPD 104
Qy      68 LTRKYGAKAYRFSISWSRIIP--LGRDLDPVNEEGIEFYSKIDALLRRTGTPWTLVYH 125
Db      105 LIAKLGDAAYRFSISWSRIIPDGLTK--INDEGITFYNNIINGLLERGIQPVYTLVYH 161
Qy      126 DLFOALHDRYGMVLNVEEYQDLFEERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGSN 185
Db      162 DLPLHLHESGWLKNTKIIEY-FAVYADTGFASGDRVQNMWITINXPMIOAIYGYATGSN 220
Qy      186 AGRSSINXSTEGNTATEPWLAKQAQIMSHARAAYVSRDPSPQKQIGISLNGDYE 245
Db      221 AGRRR-----NSLIEPYLAHHQILAAHAAVSIYRSKYKQKGGYGVVDCEMAE 272
Qy      246 PMSNERDKEAERMEFHIGFANPIFLKDYPSMKKQGLGERLPLADPADAI-LNA 304
Db      273 A-NSDKEDSAARRLDFOLGWLHPLLY-GDPEYMRRLGQQLPKFSEEDKILLNA 330
Qy      305 GENDFYGMNYTTSQFARHLDPVETDYLGAIHEHQ--ENKQSPVGESSGLAWLSCP 361
Db      331 --LDIFGLNHTYSRFISHVTECAEENHY--KVQEMERIVVEGQALIGEKASMLYVVP 387
Qy      362 DMFRKHLARYGGLYGRKPIYITENGCCPCGEBENMTCEAVNDPFRIRYFDSHLSISKAIT 421
Db      388 WGLRKILINYSOKYATPIFTENGMDDEDNDNLPLHMLDKLRYRFFKGLASVAQAI- 446
Qy      422 QDGVVVGYPFAMALLDNLWMSDGYGPRGVTFTDYTT-LKRTPKKSA 467
Db      447 KDGADVGVGFAMSLDNLFEWAQGYTKRFGLVYVDYKNGLSRHPKSSA 493

```

# RESULT 5

```

US-10-425-114-55300
; Sequence 55300, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingtong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabacka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 55300
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMFL02220131H02_FLI.pep
US-10-425-114-55300

```

Query Match 37.8%; Score 999.5; DB 12; Length 515;

Best Local Similarity 44.4%; Pred. No. 7.5e-93; Matches 212; Conservative 74; Mismatches 163; Indels 29; Gaps 11;

```

Qy      6 ALPNDPEWGFATAAYOIEGAVEGGRGPSIMDTYCHLEPSR-TNGANGVACDHRYHRYDE 64
Db      39 SPFAGTFGASSAYOIEGAVEGGRGPSIMDTYCHLEPSR-TNGANGVACDHRYHRYDE 98
Qy      65 DFLRLTKYGAAYRFSISWSRIIP--LGRDLDPVNEEGIEFYSKIDALLRRTGTPWTLVYH 124
Db      99 DVOIKGMNLDAYRFSISWSRIIPNGKLSGGINREGINYNINLHLOTGKLPVTLF 158
Qy      125 WDLPOALHDRYGMVLNVEEYQDLFEERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGS 184
Db      159 WDLPOALHDRYGMVLNVEEYQDLFEERYARLCFERFGDRVQNMWITINXPMIOAIYGYATGS 217
Qy      185 NABGRSSINXSTEGNTATEPWLAKQAQIMSHARAAYVSRDPSPQKQIGISLNGDYE 244
Db      218 KARGRSQGLRPSG--GTEPRVSHNILLAAHAAVSIYRSKYKQKGGYGVVDCEMAE 275
Qy      245 EPWDSNEPRDKEAERMEFHIGFANPIFLKDYPSMKKQGLGERLPLADPADAI-LNA 304
Db      276 VPY-SDASSDIEATERRALDFELGFMWPLTSGK-YPESMOLYVGRRLPERFKEE-DELVR 332
Qy      305 GENDFYGMNYTTSQFARHLDPVETDYLGAIHEHQ-----NKQSPVGE 351
Db      333 GSFDFGLNHTYTNTRK-----VATGYTDSVHHHDDLSTDPNVELGRLNGLSSPFGPV 386
Qy      352 SGLAWLSCPDMEFRKHLARYGGLYGRKPIYITENGCCPCGEBENMTCEAVNDPFRIRYFDP 410
Db      387 PGLGWLGVYPKGIRELLIRKUNPLIYITENGINELDDPRLSPSEBSLMDYRIDYH 446
Qy      411 SHLDSISKAITQDGVVVGYPFAMALLDNLWMSDGYGPRGVTFTDY-TTKRTPKKSA 467
Db      447 RHLLNVDYAI-RDGVVVGYPFAMALLDNLWMSDGYGPRGVTFTDY-TTKRTPKKSA 503

```

# RESULT 6

```

US-10-424-599-243532
; Sequence 243532, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```



```

Db      103 DVQIMKMNLDAYRFSISMSRILPNGKLSGINNEGINYNNLHLEQTKGKRFVTLPH 162
Qy      125 WDLPLQALHIDRYGGLWNEEYVOLDPERYARLCFERFGDRVONWITINXPMIQAITYATGS 164
Db      163 WDLPLQALENEYKGFLS-ESTIIDFGDYAKCFEEFGDRVHMWTFENPHIFSSHGAVYGT 221
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKQIMSHARAVALVYSDRPPSQGQIGISLNDY 244
Db      222 KAGGRKSQGRAPDSG--GTEPRVSHNILLAKAKAVOLYNSYKESNGEIGITLDRMF 279
Qy      245 EPWDSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPAITPADFAILNA 304
Db      280 VPY-SDASSDIEATERALDEIGWFMELTSGK-YPESMQLYVGRRLPEFSKEE-AELVR 336
Qy      305 GETDFYGMNYYTSQFAHLDGPVETDYLGAIHENO-----NKGSPIGEE 351
Db      337 GSEDFIGLNTYTTNTAR-----VATGYTDSVHHHPDSTDPVELGLTRLNGSSSPIGPV 390
Qy      352 SGLAMLRSCPMDFPKHLARVYGLYKRP-IYITENGCCPGEENMTCEAVNDPFRIRYFD 410
Db      391 PGLQMLCVYRKGIRELLRLKLNLYNPITYITENGINELDDPTLSPEBSLMDPYRIDYH 450
Qy      411 SHLDISKAITODGVVVGYPAMALLDNLEMSDGYGRFGVFTDY-TTLKRTPKSA 467
Db      451 RHLLNVDAI-RDGVRYKGYFVMSLDCFEWSNGYIRFGILFVDHKNINLRSPKLSA 507

```

## RESULT 9

```

US-10-425-114-56976
; Sequence 56976, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 56976
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17071E02_FLI.pep
US-10-425-114-56976

```

Query Match 37.8%; Score 999; DB 12; Length 527;

Best Local Similarity 42.8%; Pred. No. 8.8e-93;

Matches 208; Conservative 86; Mismatches 148; Indels 44; Gaps 15;

```

Qy      6 ALPNDPEWGPATAYQIEGAVKESGRGSPSTWDTYCHLEPRTNGANGDVACHYHRDED 65
Db      64 SFPKGFVFGTATAYQVEGAASLTNGRGPSTWDSFAHVGPIAGNQNDVAVDQYHAKED 123
Qy      66 FDLITTKGAKAYRFSISMSRIIPLG-GRDPVNEEGIEFYSKLDALLRKGITPWWLYH 124
Db      124 VDLMKSLNPAFYRFSISMSRIIPDGEK---VNEGVAAYNNLINVLLQGMTPYINLYH 180
Qy      125 WDLPLQALHIDRYGGLWNEEYVOLDPERYARLCFERFGDRVONWITINXPMIQAITYATGS 184
Db      181 YDLPLALEKKYGGMLSAKMDL-FTDYADFCFTYGRVHKMFTFENPRITALLGYDTGS 239
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKQIMSHARAVALVYSDRPPSQGQIGISLNDY 244
Db      240 NPORCT--RCAAGGNSATEPYIVAHNFLAHATAVAARVRYTKYQAAGKGVIGVLDENWY 297

```

```

Qy      245 EPWDSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPAITPADFAILNA 304
Db      298 EAL-TNSPDDQAAQAQARDPHICWFDPL-INGHYFOIQMDLVKERLPRTP-EQAKLVK 354
Qy      305 GETDFYGMNYYTSQF--ARHLDGPV-----ETDYLGAIHENOENKGSPIGESSGLA 355
Db      355 GSADYIGINEYTSYSSKQGLVQLAPSSYSADWQVYVFA-----RNGKPIGPANSK 407
Qy      356 WLRSCPMDFPKHLARVYGL-----YKRP-IYITENGCCPGEENMTCEAVNDPFRIR 407
Db      408 WLYIAP-----TGMGCVNLYKEKGNPTIYTENGNDQPG--NLTRDQYLRDATTR 458
Qy      408 YPDSHLSISKAITODGVVVGYPAMALLDNLEMSDGYGRFGVFTDYTLKRTPKSA 467
Db      459 FYRSYIGQKKALIDQ-GANVAGYFAWSLIDNFEMLAGYSSKFGIYVDFTELRHHPKASA 517
Qy      468 LVYKDM 473
Db      518 YWFRDM 523

```

## RESULT 10

```

US-10-425-114-70165
; Sequence 70165, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 70165
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01_FLI.pep
US-10-425-114-70165

```

Query Match 37.8%; Score 999; DB 12; Length 527;

Best Local Similarity 42.8%; Pred. No. 8.8e-93;

Matches 208; Conservative 86; Mismatches 148; Indels 44; Gaps 15;

```

Qy      6 ALPNDPEWGPATAYQIEGAVKESGRGSPSTWDTYCHLEPRTNGANGDVACHYHRDED 65
Db      64 SFPKGFVFGTATAYQVEGAASLTNGRGPSTWDSFAHVGPIAGNQNDVAVDQYHAKED 123
Qy      66 FDLITTKGAKAYRFSISMSRIIPLG-GRDPVNEEGIEFYSKLDALLRKGITPWWLYH 124
Db      124 VDLMKSLNPAFYRFSISMSRIIPDGEK---VNEGVAAYNNLINVLLQGMTPYINLYH 180
Qy      125 WDLPLQALHIDRYGGLWNEEYVOLDPERYARLCFERFGDRVONWITINXPMIQAITYATGS 184
Db      181 YDLPLALEKKYGGMLSAKMDL-FTDYADFCFTYGRVHKMFTFENPRITALLGYDTGS 239
Qy      185 NAGRSSINHGSTEGRNATEPMLAGKQIMSHARAVALVYSDRPPSQGQIGISLNDY 244
Db      240 NPORCT--RCAAGGNSATEPYIVAHNFLAHATAVAARVRYTKYQAAGKGVIGVLDENWY 297
Qy      245 EPWDSNEPRDKEAERMEFHIGFANPIFLKKDYPSMKKQIGERLPAITPADFAILNA 304
Db      298 EAL-TNSPDDQAAQAQARDPHICWFDPL-INGHYFOIQMDLVKERLPRTP-EQAKLVK 354
Qy      305 GETDFYGMNYYTSQF--ARHLDGPV-----ETDYLGAIHENOENKGSPIGESSGLA 355
Db      355 GSADYIGINEYTSYSSKQGLVQLAPSSYSADWQVYVFA-----RNGKPIGPANSK 407

```

[illegible]

APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaka, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ. ID NOS: 73128  
SEQ. ID NO 55483  
LENGTH: 552  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-GMFLMINOY028C03\_FLI.pep  
US-10-425-114-55483

Query Match 37.7%; Score 997; DB 12; Length 552;

Best Local Similarity 44.4%; Pred. No. 1,5e-92;

Matches 212; Conservative 78; Mismatches 169; Indels 18; Gaps 12;

QY 6 ALNDPEWGFATAVQIEGAVKEGGRPSIMDTYCHLEPSR-TNGANGDVACDHYRDE 64  
DB 72 SPDPDFIFGAGSSSYOPEGANEGRGSLIMDTFTHKYPEKIQDKSGDVAIDAYHRYKE 131  
QY 65 DFDLLKRYGAKARFSLMSRRIPLGGRDLPVNEEGIEFYSKIDALLRGITPWTLYH 124  
DB 132 DVIVIDMDNLDSYRFSISWSRILPKGLSRGINOEGIDYNNMLINELVANGIPLVTLFH 191  
QY 125 WDLPOALHRYGMLNVEEYQDPERYARLCFERFGDRVQNMITINXPMIOAIYVATGS 184  
DB 192 WDLPOSLDEBYGFLSRIYK-DPRDYAELCFKEFGGRVYKWTLINEPMSYQHGANGG 250  
QY 185 NAFGRSS-INXSTEGNTATEPMLAGKAOIMSHARAAYVSRDFRPSQKQIGISLNGD 242  
DB 251 MAFGRCSAMVNPCTGSGSTGEPYLVTHYQLAHAAVAVYKTKYQVSGKLIGITLVAN 310  
QY 243 YVPMNSNEPRDKEAERMEFHIGWPNRIFLKQVPESMKQJGRLPALPADAIL 302  
DB 311 WYLP-SNTYADKATERALIDFMGFMDPL-TSGDYPKIMRSIVRRLKFTTEOSKIL 368  
QY 303 NAGETPFYGMNYYTSQFAR--HLDGPVE--TDYLGAIEHQBENKDSPVGEESGLAWL 357  
DB 369 -IGSPFIFGANTYSSRYASAPHLNARNRYVTD---SLVTPFERGRCKIIGIKIASDWL 424  
QY 358 RSCPDMPFRKHLARVYGLYKRP-IYITENGCPCEENMTCEAVNDPFRIRYFDSHLDST 416  
DB 425 YVCPRGILDLTLTKERYNPLIYITENGINERDETLSEESILDTFRIDYHRLFYL 484  
QY 417 SKAITDDGVVYKGFAMALLDNLEMSDYGPRPGVTFTDY-TTLKRTPKKSALVLKD 472  
DB 485 RSAI-RHGVNVKGYIWSLFDNEFWSGTYRGMILVDYKNMLKRYHKLSTAIWKN 540

## RESULT 14

US-10-424-599-147332

Sequence 147332, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J

APPLICANT: Kovalic, David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ. ID NOS: 285684

SEQ. ID NO 147332

LENGTH: 503

TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10405C.1.pep  
US-10-424-599-147332

Query Match 37.7%; Score 996.5; DB 12; Length 503;

Best Local Similarity 42.6%; Pred. No. 1,5e-92;

Matches 201; Conservative 83; Mismatches 163; Indels 25; Gaps 11;

QY 8 PNDPEWGFATAVQIEGAVKEGGRPSIMDTYCHLEPSRNGANGDVACDHYRDEDFD 67  
DB 32 PNGFVFETASAFQYGAVALDEKRGPSVMDTFSHTEGKIIDFENADVAVDQYHRYEDIQ 91  
QY 68 LTRKYGAKARFSLMSRRIPLGGRDLPVNEEGIEFYSKIDALLRGITPWTLYHMDL 127  
DB 92 LMKDMGMDARFSLMSRRIPLGNG--YGOINQACVDHNLKILNMLAKGIEPVTLYHMDL 149  
QY 128 POALHRYGMLNVEEYQDPERYARLCFERFGDRVQNMITINXPMIOAIYVATGSNAP 187  
DB 150 POALHRYGMLNVEEYQDPERYARLCFERFGDRVQNMITINXPMIOAIYVATGSNAP 208  
QY 188 GRSSINXK-STEGNTATEPMLAGKAOIMSHARAAYVSRDFRPSQKQIGISLNGDYE 245  
DB 209 GRSSINXK-STEGNTATEPMLAGKAOIMSHARAAYVSRDFRPSQKQIGISLNGDYE 268  
QY 246 PMSNEPRDKEAERMEFHIGWPNRIFLKQVPESMKQJGRLPALPADAILNAG 305  
DB 269 PL-TNTEBEDIDAAQKADPQGLFMDLWPF-GDYPSSMRKRVSRLEKTSQSEALVKG-G 325  
QY 306 ETDYGMNYYTSQFARHLDGPVE--TDYLGAIEHQBENKDSPVGEESGLAWL 356  
DB 326 SLDFVGINHTTYARD-----NSTNLIGTLHLDSTADSGAVLPLFNGTKAISERASSIW 380  
QY 357 LSCPDMPFRKHLARVYGLYKRP-IYITENGCPCEENMTCEAVNDPFRIRYFDSHLDST 415  
DB 381 LYVPMNSNEPRDKEAERMEFHIGWPNRIFLKQVPESMKQJGRLPALPADAILNAG 440  
QY 416 ISKAITDDGVVYKGFAMALLDNLEMSDYGPRPGVTFTDY-TTLKRTPKKS 466  
DB 441 LLSAI-KDGCNVKGYIWSLFDNEFWSGTYRGMILVDYKNMLKRYHKLSTAIWKN 491

## RESULT 15

US-10-424-599-219958

Sequence 219958, Application US/10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J

APPLICANT: Kovalic, David K

APPLICANT: Zhou Yihua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ. ID NOS: 285684

SEQ. ID NO 219958

LENGTH: 510

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: unsure

LOCATION: (11).(510)

OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_4064C.1.pep  
US-10-424-599-219958

Query Match 37.3%; Score 986; DB 12; Length 510;

Best Local Similarity 44.3%; Pred. No. 1,8e-91;

Matches 213; Conservative 77; Mismatches 167; Indels 24; Gaps 12;

```

QY 2 PESIAL-----PNDFENGFAITAAVOIGAVKEGGRPSIMDTYCHLEPSRTNGA-NGDVA 55
Db 31 PEIASLNRNSFPPTGFIPTASASAYQYEGANEGGRPSIMDTFTKYPDKIKORDSGDVA 90
QY 56 CDHHRVDEDEPDLLTXGAKAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKLIIDALIRG 115
Db 91 IDSYHRYKEDVGIMKDMNLDAYRPSISMSRIIPKGLSGGINQBGIDYNNLINELLANG 150
QY 116 ITPWVTLYHMDLPQALHDRVGMNLVEBVOLDPERYARLCEPFGDRVQNMVITINXPMIQ 175
Db 151 LKPEVTILFHMDLPGSLDEBYGFLSPRIVK-DFQDYADLCPEKFGDRVKHMITLINEPWSY 209
QY 176 AIYGVATGSNAPGRSS--INKHSTEGNTATEPWLAKGAQIMSHARAVAYSRDPRPSQKG 233
Db 210 SQHGVAATGEMAPGRCSAMNPNCGDSDATEPYLVSHHQLAHAAVAVHYKTKYQTFONG 269
QY 234 QIGISLNDYVEPMDSNBPRDKEAERMEFPIGFANPIFLKKDYPSMKKOLGERLPA 293
Db 270 LIGITLNVNMVYVPSDNK-LDHKATERAIDFQYGFMDPL-TTGDYPSKSMFLVBARLPK 327
QY 294 LTPADFAILNAGETDFYGMNYTSGQFARHLDPVPEPTDYLGAIHENKDKDSPVGEESG 353
Db 328 FTKEQSKLL-IDSFDFIGINYSASAYS--DAPOLSNAKISYLTDSLNSXSAS----- 378
QY 354 LAMLRSCPDMPKHLARVYGLYKRP-IYITENGCPCPGEENMTCEAVNDPFRIRYFDSH 412
Db 379 -NMLYVPRGRFDVLLYTKKKYNNPLIYITENGINEYDSSLSLESLLDIYRIDYHRH 437
QY 413 LDISISKAITODGVVYGYFAMALLDNLWSDGYGPRFGVFTDY-TTLKRTPKKSALVLK 471
Db 438 LFYLOEAL-KRGVNVKGYFAMSLDNPFMHIGYTVRFGMNFIIDYKNDLKRYSKLSALWFK 496
QY 472 D 472
Db 497 D 497

```

Search completed: March 25, 2004, 16:02:25  
 Job time : 47 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 25, 2004, 15:55:12 ; Search time 182 Seconds  
(without alignments)  
2595.659 Million cell updates/sec

Title: US-10-026-140-2  
Perfect score: 2644  
Sequence: 1 MPESLALPDMFEGFATPAAY.....KSALVLKDMFAARQVVA 484

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Pending\_Patents\_AA\_Main: \*  
1: /cgn2\_6/prodata/2/paa/US06\_PCTUS\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/prodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/prodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/prodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/prodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/prodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/prodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/prodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/prodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/prodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/prodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/prodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/prodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/prodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/prodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/prodata/2/paa/US097A\_COMB.pep.\*  
22: /cgn2\_6/prodata/2/paa/US097B\_COMB.pep.\*  
23: /cgn2\_6/prodata/2/paa/US098\_COMB.pep.\*  
24: /cgn2\_6/prodata/2/paa/US099\_COMB.pep.\*  
25: /cgn2\_6/prodata/2/paa/US099B\_COMB.pep.\*  
26: /cgn2\_6/prodata/2/paa/US100\_COMB.pep.\*  
27: /cgn2\_6/prodata/2/paa/US101\_COMB.pep.\*  
28: /cgn2\_6/prodata/2/paa/US102\_COMB.pep.\*  
29: /cgn2\_6/prodata/2/paa/US103\_COMB.pep.\*  
30: /cgn2\_6/prodata/2/paa/US104\_COMB.pep.\*  
31: /cgn2\_6/prodata/2/paa/US106\_COMB.pep.\*  
32: /cgn2\_6/prodata/2/paa/US107\_COMB.pep.\*  
33: /cgn2\_6/prodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	2642	99.9	484	1	PCT-US02-34674-2	Sequence 2, Appli
2	2642	99.9	484	1	PCT-US02-34764-2	Sequence 2, Appli
3	2642	99.9	484	26	US-10-026-140-2	Sequence 2, Appli
4	1303.5	49.3	476	22	US-09-791-537-52753	Sequence 52753, A
5	1270	48.0	466	22	US-09-791-537-52753	Sequence 52755, A
6	1210	45.8	460	29	US-10-369-493-13405	Sequence 13405, A
7	1210	45.5	490	33	US-60-360-039-31405	Sequence 13405, A
8	1204	45.5	524	18	US-09-417-507-31496	Sequence 31496, A
9	1058.5	40.0	491	30	US-10-424-599-169310	Sequence 169310, A
10	1058.5	40.0	511	28	US-10-219-999-35761	Sequence 35761, A
11	1058.5	40.0	511	30	US-10-425-114-43168	Sequence 43168, A
12	1058.5	40.0	511	33	US-10-425-114-43168	Sequence 43168, A
13	1058.5	40.0	511	33	US-60-324-109-21315	Sequence 21315, A
14	1038	39.3	521	30	US-10-437-963-199890	Sequence 199890, A
15	1037	39.2	514	19	US-09-513-996A-34109	Sequence 34109, A
16	1037	39.2	516	19	US-09-513-996A-34108	Sequence 34108, A
17	1037	39.2	523	19	US-09-513-996A-34107	Sequence 34107, A
18	1025.5	38.8	534	24	US-09-935-625-12100	Sequence 12100, A
19	1020	38.6	515	30	US-10-437-963-144610	Sequence 144610, A
20	1013	38.3	537	22	US-09-791-537-13983	Sequence 13983, A
21	1006	38.0	510	30	US-10-437-963-202947	Sequence 202947, A
22	1003	37.9	577	22	US-09-791-537-80861	Sequence 80861, A
23	999.5	37.8	515	28	US-10-219-999-43969	Sequence 43969, A
24	999.5	37.8	515	30	US-10-425-114-55300	Sequence 55300, A
25	999.5	37.8	515	30	US-10-425-114-55300	Sequence 55300, A
26	999.5	37.8	515	33	US-60-324-109-30599	Sequence 30599, A
27	999.5	37.8	516	30	US-10-424-599-243532	Sequence 243532, A
28	999.5	37.8	519	28	US-10-219-999-36333	Sequence 36333, A
29	999.5	37.8	519	28	US-10-219-999-47299	Sequence 47299, A
30	999.5	37.8	519	30	US-10-425-114-48403	Sequence 48403, A
31	999.5	37.8	519	30	US-10-425-114-55444	Sequence 55444, A
32	999.5	37.8	519	30	US-10-425-114A-48403	Sequence 48403, A
33	999.5	37.8	519	30	US-10-425-114A-55444	Sequence 55444, A
34	999.5	37.8	519	33	US-60-312-544-6898	Sequence 6898, A
35	999.5	37.8	519	33	US-60-324-109-31263	Sequence 31263, A
36	999	37.8	502	28	US-10-219-999-48605	Sequence 48605, A
37	999	37.8	502	28	US-10-219-999-61094	Sequence 61094, A
38	999	37.8	502	28	US-10-219-999-61151	Sequence 61151, A
39	999	37.8	502	33	US-60-324-109-27920	Sequence 27920, A
40	999	37.8	502	33	US-60-324-109-29363	Sequence 29363, A
41	999	37.8	502	33	US-60-324-109-29339	Sequence 29339, A
42	999	37.8	524	33	US-60-376-344-4	Sequence 4, Appli
43	999	37.8	527	30	US-10-425-114-56976	Sequence 56976, A
44	999	37.8	527	30	US-10-425-114-70165	Sequence 70165, A
45	999	37.8	527	30	US-10-425-114-71098	Sequence 71098, A

## ALIGNMENTS

RESULT 1  
PCT-US02-34674-2  
Sequence 2, Application PC/TUS0234674  
GENERAL INFORMATION:  
APPLICANT: Dunn-Coleman, Nigel  
APPLICANT: Goedegebuur, Frits  
APPLICANT: Ward, Michael  
TITLE OF INVENTION: BGLS Beta-Glucosidase and Nucleic Acids  
FILE REFERENCE: GC697  
CURRENT APPLICATION NUMBER: PCT/US02/34674  
CURRENT FILING DATE: 2003-03-14  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq For Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Trichoderma reesei  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1) .. (484)  
OTHER INFORMATION: Xaa = Any Amino Acid

PCT-US02-34674-2

Query Match 99.9%; Score 2642; DB 1; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.5e-276;  
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60
QY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
QY 121 TLYHMDLPQALHMDRYGGMNLVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAIYGY 180
DB 121 TLYHMDLPQALHMDRYGGMNLVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAIYGY 180
QY 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAOIMSHARAAYVSRDPRPSOKGIGISLN 240
DB 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAOIMSHARAAYVSRDPRPSOKGIGISLN 240
QY 241 GDIYEPWDSNEPRDKEAERMEFHIGWFANPIFLKKDYPESMKQOLGERLPALTADFA 300
DB 241 GDIYEPWDSNEPRDKEAERMEFHIGWFANPIFLKKDYPESMKQOLGERLPALTADFA 300
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHOKDGSFVGEESGLAMLRSC 360
DB 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHOKDGSFVGEESGLAMLRSC 360
QY 361 PDMEFKHLARVYGLYKXPYITTEGCPCEENMTCEAVNDPFRIRYFDSHLSISKAI 420
DB 361 PDMEFKHLARVYGLYKXPYITTEGCPCEENMTCEAVNDPFRIRYFDSHLSISKAI 420
QY 421 TODGVVVKGYFAMALLDNLEMSDGYGPRFVGTFTDYTLTKRTPKKSALVTKDMFAAROV 480
DB 421 TODGVVVKGYFAMALLDNLEMSDGYGPRFVGTFTDYTLTKRTPKKSALVTKDMFAAROV 480
QY 481 KVAA 484
DB 481 KVAA 484
```

RESULT 2  
PCT-US02-34764-2  
Sequence 2, Application PC/TUS0234764  
GENERAL INFORMATION:  
APPLICANT: Dunn-Coleman, Nigel  
APPLICANT: Goedegebuurt, Frits  
APPLICANT: Ward, Michael  
APPLICANT: Yao, Jian  
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids  
FILE REFERENCE: GC697  
CURRENT APPLICATION NUMBER: PCT/US02/34764  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Trichoderma reesei  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(484)  
OTHER INFORMATION: Xaa = Any Amino Acid  
PCT-US02-34764-2

Query Match 99.9%; Score 2642; DB 1; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.5e-276;  
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60

```
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60
QY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120
QY 121 TLYHMDLPQALHMDRYGGMNLVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAIYGY 180
DB 121 TLYHMDLPQALHMDRYGGMNLVEEVDLPERYARLCFERFGDRVQNMWITINXPMIOAIYGY 180
QY 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAOIMSHARAAYVSRDPRPSOKGIGISLN 240
DB 181 ATGSNAPGRSSINKHSTEGNTATEPMLAGKAOIMSHARAAYVSRDPRPSOKGIGISLN 240
QY 241 GDIYEPWDSNEPRDKEAERMEFHIGWFANPIFLKKDYPESMKQOLGERLPALTADFA 300
DB 241 GDIYEPWDSNEPRDKEAERMEFHIGWFANPIFLKKDYPESMKQOLGERLPALTADFA 300
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHOKDGSFVGEESGLAMLRSC 360
DB 301 ILNAGETDFYGMNYYTSQFARHLDPVPETDYLGAIHEHOKDGSFVGEESGLAMLRSC 360
QY 361 PDMEFKHLARVYGLYKXPYITTEGCPCEENMTCEAVNDPFRIRYFDSHLSISKAI 420
DB 361 PDMEFKHLARVYGLYKXPYITTEGCPCEENMTCEAVNDPFRIRYFDSHLSISKAI 420
QY 421 TODGVVVKGYFAMALLDNLEMSDGYGPRFVGTFTDYTLTKRTPKKSALVTKDMFAAROV 480
DB 421 TODGVVVKGYFAMALLDNLEMSDGYGPRFVGTFTDYTLTKRTPKKSALVTKDMFAAROV 480
QY 481 KVAA 484
DB 481 KVAA 484
```

RESULT 3  
US-10-026-140-2  
Sequence 2, Application US/10026140  
GENERAL INFORMATION:  
APPLICANT: Dunn-Coleman, Nigel  
APPLICANT: Goedegebuurt, Frits  
APPLICANT: Ward, Michael  
APPLICANT: Yao, Jian  
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids  
FILE REFERENCE: GC697  
CURRENT APPLICATION NUMBER: US/10/026,140  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 484  
TYPE: PRT  
ORGANISM: Trichoderma reesei  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)...(484)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-026-140-2

Query Match 99.9%; Score 2642; DB 26; Length 484;  
Best Local Similarity 100.0%; Pred. No. 1.5e-276;  
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60  
DB 1 MPESLALPNDPEWGFATTAAYOIEGAVKEGGRGPSIMDTYCHLEPRTNGANGDVACDHYH 60  
QY 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120  
DB 61 RYDEDFDLTKYGAAYRFSLSMSRIIPLGRLDPVNEEGIEFYSKIDALLRGGITPMV 120

QY 121 TLVHMDLPQALHRYGGMILNVEEVLDFEYARLCEFERFGDRVQNMWTTXNPMIOAIYGY 180  
DB 121 TLVHMDLPQALHRYGGMILNVEEVLDFEYARLCEFERFGDRVQNMWTTXNPMIOAIYGY 180  
QY 181 ATGSNAFGRSSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSDRFRPQCKQIGISLN 240  
DB 181 ATGSNAFGRSSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSDRFRPQCKQIGISLN 240  
QY 241 GDVYEPMDSEPRDKEAERMEFHIGWFPANPIFLKQDYPESMKQGLGEPLPALTTPADFA 300  
DB 241 GDVYEPMDSEPRDKEAERMEFHIGWFPANPIFLKQDYPESMKQGLGEPLPALTTPADFA 300  
QY 301 ILNAGETDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSC 360  
DB 301 ILNAGETDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSC 360  
QY 361 PDMFRKHLARVGLYKGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAI 420  
DB 361 PDMFRKHLARVGLYKGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAI 420  
QY 421 TQGVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 480  
DB 421 TQGVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 480  
QY 481 KVAA 484  
DB 481 KVAA 484

RESULT 4  
US-09-791-537-52753  
; Sequence 52753, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; TITLE OF INVENTION: METHODS OF USE THEROP  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 52753  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Humicola grisea var  
US-09-791-537-52753

Query Match 49.3%; Score 1303.5; DB 22; Length 476;  
Best Local Similarity 51.7%; Pred. No. 2.9e-131;  
Matches 244; Conservative 79; Mismatches 144; Indels 5; Gaps 5;

QY 5 LALPNDPEWGFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 64  
DB 1 MSLPDPKMGFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 60  
QY 65 DFDLLTKYKAKAYRSLSMRRIIPLGRLDPVNEEGIEFYSKLDALLRGITPWTLYH 124  
DB 61 DIALLEKELGANSYFSSISWSRIIPLGGRNDPINKGIDHYKFPVDDIEAGITPFIPLFH 120  
QY 125 WDLPOLAHRYGGMILNVEEVLDFEYARLCEFERFGDRVQNMWTTXNPMIOAIYGYTGS 184  
DB 121 WDLPOLAHRYGGMILNVEEVLDFEYARLCEFERFGDRVQNMWTTXNPMIOAIYGYTGS 179  
QY 185 NABGRSSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSDRFRPQCKQIGISLANDY 244  
DB 180 FAGHTSDRSKSPVGDARREPMIIGHNIIIAHARAVAVYSDRFRPQCKQIGISLANDY 239  
QY 245 EPMDSNEPRDKEAERMEFHIGWFPANPIFLKQDYPESMKQGLGEPLPALTTPADFA 304  
DB 240 LPMDEPDPADIEACDRKIEFAISWADPIYFGK-YPSMRKQGLDRLEPFTPEEVALVK- 297

QY 305 GETDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSCDPMF 364  
DB 298 GSDNDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSCDPMF 357  
QY 365 RKLHARVGLYKGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAITOP 423  
DB 358 RDLNLMKSRKYGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAITOP 417  
QY 424 GVVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 474  
DB 418 GCVNRGILAMSLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 469

RESULT 5  
US-09-791-537-52755  
; Sequence 52755, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Biomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB  
; TITLE OF INVENTION: METHODS OF USE THEROP  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 52755  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Hypocrea jecorina  
US-09-791-537-52755

Query Match 48.0%; Score 1270; DB 22; Length 466;  
Best Local Similarity 52.7%; Pred. No. 1.2e-127;  
Matches 246; Conservative 65; Mismatches 142; Indels 16; Gaps 7;

QY 7 LPLNDFEWFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 66  
DB 2 LPLNDFEWFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 61  
QY 67 DLTLYKAKAYRSLSMRRIIPLGRLDPVNEEGIEFYSKLDALLRGITPWTLYHMD 126  
DB 62 ALKSLGAKSRYFSSISWSRIIPGGRNDPINKGIDHYKFPVDDIEAGITPFIPLFHMD 121  
QY 127 LPLNDFEWFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 186  
DB 122 LPLNDFEWFATAVOIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDE 180  
QY 187 PGRSSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSDRFRP-SQKQIGISLANDY 245  
DB 181 PGRSSINKHSTEGNTATEPWLAKQAQIMSHARAVAVYSDRFRP-SQKQIGISLANDY 230  
QY 246 PWDNEPRDKEAERMEFHIGWFPANPIFLKQDYPESMKQGLGEPLPALTTPADFA 305  
DB 231 PWDNEPRDKEAERMEFHIGWFPANPIFLKQDYPESMKQGLGEPLPALTTPADFA 288  
QY 306 ETPDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSCDPMF 365  
DB 289 SNDFYGMNYTTSQFARHLDPVPEYDYLGAIEHONKDGSPVGESEGLAWLRSCDPMF 348  
QY 366 KHLARVGLYKGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAITOP 424  
DB 349 DFLWISKRYGPRIYITTEGCPGCEENMTCEAVNDPRIRYFDSHLSISKAITOP 408  
QY 425 VVVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 474  
DB 409 VVVVVKGYFAMALLDNLEMSDGYGPRFGVTFTDYLTKRTPKKSALVLDMPAARQV 459

RESULT 6  
US-10-369-493-13405  
; Sequence 13405, Application US/10369493

```
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Hankle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Chen, Xianfeng
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)B
/ CURRENT APPLICATION NUMBER: US/10/369,493
/ CURRENT FILING DATE: 2003-02-28
/ PRIOR APPLICATION NUMBER: US 60/360,039
/ PRIOR FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 13405
/ LENGTH: 490
/ TYPE: PRT
/ ORGANISM: Aspergillus nidulans
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(490)
/ OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-13405
```

```
Query Match          45.8%; Score 1210; DB 29; Length 490;
Best Local Similarity 47.3%; Pred. No. 4.3e-121;
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;
```

```
QY 7 LPNDEWGFATAAVQIEGAVKGGKPSITWDYCHLEPSRTNGANGDVACHYHYRDEDF 66
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1 LRNDFHGVATTAAGVAGAMNKDGKPSITWDTGHTPGKVKONSADVAVRFYDFRBDV 60
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 67 DLITVYGAARFSLWSRIIPLGRLDPVNEEGIEFYSKLIDALLRGGITPWTLYHMD 126
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 61 ALMKSYGVNAVYFSLWSRIIPLGADDPVNEGGIKYYQDLVDLNNIGITPVTLLFHM 120
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 127 LPQALHDRVGMNLNEVQDLDFERYARLCFERFGDRVOMTITNKPWIAIYGATGNSA 186
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 121 VPQALHEDRYGMNLNQRPIPDVRYARVCFERLGPVXHMWTFNEPGVYSLAGYAGVHA 180
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 187 PGRSSINRGSTEGNTATEPMLAGKAQIMSHARAVALVSRDPSPQKQIGISLNGDYEP 246
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 181 PARSSRELNEEGSDSTEPIVGHTKLVTHGVSKLYREVFQPOQKGTIGITLHGWSBP 240
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 247 WDSNEPRKEAARERMEFHIGWPNPIFLKKDYPESMKQIGERLPAITPADFAILNAGE 306
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 241 WDEDDPRQEAERAREFEIAMPDPPLKTDYDPAQMLGDRLPRTTPESSKLV-LGS 299
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 307 TDFYGMNNTYTSQFARHLDPVPETDYLGAIHGHQENKQSPVGEESGLAMLRSCPDMPFK 366
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 300 SEFYGMNSTYTFVQHKDTPPDINDHKGNVIVHDNNSKGVSRGEESDTPWLRATPTGMRK 359
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 367 HLAARYGLGKPIYITENGCPCPEENMTCEAVNDPRIRYFDSHLSISKAITQDGYV 426
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 360 LLMWIMNRYHVPYITENGTAKGETAPT-PEVLIDTFRRMRFFBGVGLARAKKEGVD 418
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 427 VKGYFAMALLDN-----LEWSDGYGPRFGVTFPTDYTLTKRT--PKKSAL 468
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 419 IRSYFAWTFITDWMGMFKPFPPLSDXKRVEMAAAGYTRFCFTIDFDSPMKTRYPKQSA 478
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 469 VLKDMF 474
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 479 YLKALF 484
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
RESULT 7
/ Sequence 13405, Application US/60360039
/ GENERAL INFORMATION:
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Chen, Xianfeng
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hankle, Gregory J.
```

```
/ APPLICANT: Slater, Steven C.
/ TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
/ FILE REFERENCE: 38-10(52052)A
/ CURRENT APPLICATION NUMBER: US/60/360,039
/ CURRENT FILING DATE: 2002-02-21
/ NUMBER OF SEQ ID NOS: 47374
/ SEQ ID NO 13405
/ LENGTH: 490
/ TYPE: PRT
/ ORGANISM: Aspergillus nidulans
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(490)
/ OTHER INFORMATION: unsure at all Xaa locations
US-60-360-039-13405
```

```
Query Match          45.8%; Score 1210; DB 33; Length 490;
Best Local Similarity 47.3%; Pred. No. 4.3e-121;
Matches 230; Conservative 82; Mismatches 154; Indels 20; Gaps 4;
```

```
QY 7 LPNDEWGFATAAVQIEGAVKGGKPSITWDYCHLEPSRTNGANGDVACHYHYRDEDF 66
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1 LRNDFHGVATTAAGVAGAMNKDGKPSITWDTGHTPGKVKONSADVAVRFYDFRBDV 60
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 67 DLITVYGAARFSLWSRIIPLGRLDPVNEEGIEFYSKLIDALLRGGITPWTLYHMD 126
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 61 ALMKSYGVNAVYFSLWSRIIPLGADDPVNEGGIKYYQDLVDLNNIGITPVTLLFHM 120
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 127 LPQALHDRVGMNLNEVQDLDFERYARLCFERFGDRVOMTITNKPWIAIYGATGNSA 186
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 121 VPQALHEDRYGMNLNQRPIPDVRYARVCFERLGPVXHMWTFNEPGVYSLAGYAGVHA 180
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 187 PGRSSINRGSTEGNTATEPMLAGKAQIMSHARAVALVSRDPSPQKQIGISLNGDYEP 246
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 181 PARSSRELNEEGSDSTEPIVGHTKLVTHGVSKLYREVFQPOQKGTIGITLHGWSBP 240
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 247 WDSNEPRKEAARERMEFHIGWPNPIFLKKDYPESMKQIGERLPAITPADFAILNAGE 306
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 241 WDEDDPRQEAERAREFEIAMPDPPLKTDYDPAQMLGDRLPRTTPESSKLV-LGS 299
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 307 TDFYGMNNTYTSQFARHLDPVPETDYLGAIHGHQENKQSPVGEESGLAMLRSCPDMPFK 366
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 300 SEFYGMNSTYTFVQHKDTPPDINDHKGNVIVHDNNSKGVSRGEESDTPWLRATPTGMRK 359
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 367 HLAARYGLGKPIYITENGCPCPEENMTCEAVNDPRIRYFDSHLSISKAITQDGYV 426
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 360 LLMWIMNRYHVPYITENGTAKGETAPT-PEVLIDTFRRMRFFBGVGLARAKKEGVD 418
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 427 VKGYFAMALLDN-----LEWSDGYGPRFGVTFPTDYTLTKRT--PKKSAL 468
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 419 IRSYFAWTFITDWMGMFKPFPPLSDXKRVEMAAAGYTRFCFTIDFDSPMKTRYPKQSA 478
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 469 VLKDMF 474
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 479 YLKALF 484
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
RESULT 8
/ Sequence 31496, Application US/09417507
/ GENERAL INFORMATION:
/ APPLICANT: KEITH G. WEINSTOCK ET AL.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
/ FILE REFERENCE: PATH99-10
/ CURRENT APPLICATION NUMBER: US/09/417,507
/ CURRENT FILING DATE: 1999-10-14
/ NUMBER OF SEQ ID NOS: 44312
/ SEQ ID NO 31496
/ LENGTH: 524
/ TYPE: PRT
/ ORGANISM: A.fumigatus
```

```

FEATURES:
; NAME/KEY: UNSURE
; LOCATION: (521)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknc
US-09-417-507-31496

```

```

Query Match      45.5%; Score 1204; DB 18; Length 524;
Best Local Similarity 46.9%; Pred. No. 2,2e-120;
Matches 229; Conservative 82; Mismatches 155; Indels 22; Gaps 5;

```

```

QY 7 LPNDFEKGFTAAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 66
DB 20 LRPDFHFGVATAAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 79
QY 67 DLTTKYGAAYRFSLSMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 126
DB 80 ALMKSYGVNAYRFSLSMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 139
QY 127 LPQALHDIRYGMVLANVEEVLQDFERYARLCEFRFGDRVQNMITINXPMWQAIYGYATGSNA 186
DB 140 TPGSLDERYGMVLANVEEVLQDFERYARLCEFRFGDRVQNMITINXPMWQAIYGYATGSNA 199
QY 187 PGRSSINKHSTEGNTATEPMLAGKQIMSHARAVAVSRDPSPQKQIGISLNGDYEP 246
DB 200 PGRSSINKHSTEGNTATEPMLAGKQIMSHARAVAVSRDPSPQKQIGISLNGDYEP 259
QY 247 WDSENPBRDKEAARMEFHIQWFPANPIFLKDYBPESMKQLGERLPALTAPADFAILLAGE 306
DB 260 WDEADLPDQAAAEAREFEIAMPADPLKGDYFASHRAQDGLPFTPEBESKLV--LGS 318
QY 307 TDEYGMNYYTSQFARHLDPVPEITDYLGAIHEHQENKDGSPVGEESGLAWLRSCPDMPFRK 366
DB 319 SEFYGMNYYTSQFARHLDPVPEITDYLGAIHEHQENKDGSPVGEESGLAWLRSCPDMPFRK 378
QY 367 HLAHYGLYKPIYITENGCPGCEENMTCEAVNDPFRIRYPSDLSH--SISKAITDDGV 425
DB 379 LLMWISRYQMPITYVENGTATAG--ETAPSPSVLNDQFRIRFEGYGVGMALARAIVKEDGI 437
QY 426 VVSGYFPMALLD-----LEWSDGYPFGVFTPTTYTLKRT--PKKS 466
DB 438 DISYFPAWTTTDMWGFPRFLSSCKNAEDGTWMAAGYTDPRGCTFDPSPEKTRYPKOS 497
QY 467 ALVLKDMF 474
DB 498 AYLLDNLF 505

```

## RESULT 9

```

US-10-424-599-169310
; Sequence 169310, Application US/10424599
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169310
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123901C.1.pcp
US-10-424-599-169310

```

```

Query Match      40.0%; Score 1058.5; DB 30; Length 491;
Best Local Similarity 46.7%; Pred. No. 1.2e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

```

QY 8 PNDPEWGFATAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 67
DB 25 PNFIFGVATSAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 84
QY 68 DLTTKYGAAYRFSLSMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 125
DB 85 LIAKLGFDARFISMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 141
QY 126 DLQALHDIRYGMVLANVEEVLQDFERYARLCEFRFGDRVQNMITINXPMWQAIYGYATGSN 185
DB 142 DLPALHIESGGLNKOIIEY--FAVYADTCFASFGDRVQNMITINXPMWQAIYGYATGSN 200
QY 186 AGRSSINKHSTEGNTATEPMLAGKQIMSHARAVAVSRDPSPQKQIGISLNGDYEP 245
DB 201 AGRRE-----NSLIEPYLAHHQIILAHAAVAVSYRSKYKQKQGGVFDCEMAE 252
QY 246 PMSNEPRDKEAARMEFHIQWFPANPIFLKDYBPESMKQLGERLPALTAPADFAILLAGE 304
DB 253 A-NSDKIEDKSAARLDFQGFHLPVLY--GDYPEVMERLQDLPKFEEDKAILNA 310
QY 305 GETDFYGMNYYTSQFARHLDPVPEITDYLGAIHEHQ--ENKDGSPVGEESGLAWLRSCP 361
DB 311 --LDFTGLNHYTRFISHVTECAENHY--KVQEMERIVEMEGQAIQEAASEMLVYP 367
QY 362 DMFRKHLARYGLYKPIYITENGCPGCEENMTCEAVNDPFRIRYPSDLSH--SISKAIT 421
DB 368 WGRKILNLYVSQYATPIFTENGMDDEDNDLPLHEMLDKLRYRFKGYLASVAQAI-- 426
QY 422 QDGVVYKGYFAMALLDNLEWSDGYPFGVFTPTDYTL--LKRTPKKA 467
DB 427 KQDADVRYGFAMSLDNLFEMWAGQYTRFGILVYVDYKNGLSRHKSSA 473

```

## RESULT 10

```

US-10-219-999-35761
; Sequence 35761, Application US/10219999
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52722)C
; CURRENT APPLICATION NUMBER: US/10/219,999
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 60/324,109
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/312,544
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 63520
; SEQ ID NO 35761
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
US-10-219-999-35761

```

```

Query Match      40.0%; Score 1058.5; DB 28; Length 511;
Best Local Similarity 46.7%; Pred. No. 1.3e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

```

```

QY 8 PNDPEWGFATAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 67
DB 45 PNFIFGVATSAVQIEGAVKEGGRGPSIWDYTCHELEPSRTNGANGVACDHRHYDEDF 104
QY 68 DLTTKYGAAYRFSLSMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 125
DB 105 LIAKLGFDARFISMSRIIP--LGGRLDPVNEEGIEFYSKILDALLRGRTTPWVTLVYHW 161
QY 126 DLQALHDIRYGMVLANVEEVLQDFERYARLCEFRFGDRVQNMITINXPMWQAIYGYATGSN 185
DB 162 DLPALHIESGGLNKOIIEY--FAVYADTCFASFGDRVQNMITINXPMWQAIYGYATGSN 220

```



```

; Sequence 21315, Application US/60324109
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingsong
; APPLICANT: Stein, Joshua
; TITLE OF INVENTION: CDNA SEQUENCES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-10(52726)B
; CURRENT APPLICATION NUMBER: US/60/324,109
; CURRENT FILING DATE: 2001-09-21
; NUMBER OF SEQ ID NOS: 31196
; SEQ ID NO 21315
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
;
US-60-324-109-21315

Query Match      40.0%; Score 1058.5; DB 33; Length 511;
Best Local Similarity 46.7%; Pred. No. 1.3e-104;
Matches 218; Conservative 77; Mismatches 147; Indels 25; Gaps 12;

QY      8 PNDPENGATAYQIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDEDF
DB      45 PPFIFIGVATSAQIIGACKEGGRPSINDAFTHTEGKLIDKSNQDVAVNHHRYMEDID
QY      68 LITKYGAKAYRFSLSRSRIIP-LGGRLDVNEEGIEFYSKIDALLRGITPWTLYHW
DB      105 LIAKLGFDAYRFSISMSRIFFPDGLGK---INDEGITFYNNIINGLIERGIQPYVTLVHW
QY      126 DLPQALHDYRGMLNVEVQDLDFERYARLCEFERGDRVQNMWITINXPMWIAIYGATGSN
DB      162 DLPRLHESMGMLNKOILLEY-PAVYADTCFASFGGRVKWITINBPLOAVNGYDVALF
QY      186 AGRSSINHGTEGNTATPEPMLAGKAQIMSHARAVAVYSRDPSPQKQIGISLNGDY
DB      221 AGRRE-----NLIIEPYLAHHQILAHAAVSIYRSKYDKKQCGVGFVDDCEMAE
QY      246 PNDSPNEPRDKEAERMEFIHGFANPIFLKKDYPSMKKQIGERLPALTPADPAI-LNA
DB      273 A-NSDKIEBKSAARRLDFOLGFWFLHPLY-GDYPERMRRLDQLPKFEEDPKILLNA
QY      305 GEDDFGKMYNYSQFARHLDPVPEPTDYLGAIHQ---ENKDGSPVGEESGLAMLRSCP
DB      331 --LDFTGLNHYTSRFLSHYTECAEENHYX-KVQEMERIVEMEGGOAIGEKASAEMLYVP
QY      362 DMFRKHLAVYGLYGRPIYITENGCCPGCEENMTCEAVNDPRIRYFDSHLSISKAIT
DB      388 WGRKRLINVSQKATPIPTTEGMDDEDNDNLPLEHMLDDKLRVYFKGYLVAQAII-446
QY      422 QDGVVVKGFAMALLDNLEWSDGYGRPGVYFTDYTT-LKRTPKKA 467
DB      447 KQGAADVGRYFAMSLDNFEMAQGYTKRFGIVYDYKNGKLSRHKSSA 493

RESULT 14
US-10-437-963-199890
; Sequence 199890, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

```

```

; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 199890
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_95413C.1.pcp
US-10-437-963-199890

Query Match      39.3%; Score 1038; DB 30; Length 521;
Best Local Similarity 44.7%; Pred. No. 2.2e-102;
Matches 212; Conservative 80; Mismatches 158; Indels 24; Gaps 10;

QY      6 ALPNDPENGATAYQIEGAVKEGGRPSIMDTYCHLEPSRTNGANGVACDHYHRYDEDF
DB      49 SPEEGVFPGTASAAVSDGKEDGQITWDTWTFATFGKITDPSNADVAVDYHHEED
QY      66 FDLITRYGAKAYRFSLSRSRIIP-LGGRLDVNEEGIEFYSKIDALLRGITPWTLYHW
DB      109 IQLMADMGMDAYRFSIAMSRIYNG--VQVNOAGIDHNKILIDALLANGIQPYVTLVHW
QY      126 DLPQALHDYRGMLNVEVQDLDFERYARLCEFERGDRVQNMWITINXPMWIAIYGATGSN
DB      167 DLPQALHDYRGMLNVEVQDLDFERYARLCEFERGDRVQNMWITINXPMWIAIYGATGSN
QY      186 AGRSSINHG--STEGNTATPEPMLAGKAQIMSHARAVAVYSRDPSPQKQIGISLNGDY
DB      226 AGRCSVLLHLYCKAANSSTEPYVVAHFLTLAAASIRTKYKATQNGQIDAIADPVMW
QY      244 YEPWDSNEPRDKEAERMEFIHGFANPIFLKKDYPSMKKQIGERLPALTPADPAI-LN
DB      286 FEPW-SNTTIDIEAAGRAGEFOLGWFADPFF-GDYPMARAVGRRLPFT-ADEAAV
QY      304 AGEDDFGKMYNYSQFARHLDPVPEPTDYLGAIHQ-----ENKDGSPVGEESGL
DB      343 KGALEDVGVNHYTYTTRH-----NNTNIIIGTLNNTLADTGVSILPFGKXKPIGRANS
QY      355 AMWRSCPMFRKHLAVYGLYGRPIYITENGCCPGCEENMTCEAVNDPRIRYFDSHL
DB      398 IWLIVYPRGRSILMYNYSKRYNSPVYITENGDDSNPSTISKDLKOSKRLKYNHNDYL
QY      414 DSISKAITDGVVVKGFAMALLDNLEWSDGYGRPGVYFTDY-TLKRTPKKS 466
DB      458 TNLASIKEDGCVRYGFAMSLDNFEMAAGYSRGLYFVYKDLKRYPKKS 511

RESULT 15
US-09-513-996A-34109
; Sequence 34109, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. Alexandrov et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513,996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 34109
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..514
; OTHER INFORMATION: any n or Xaa = unknown
; OTHER INFORMATION: Location 1..514 / Ceres Seq. ID 1715035
US-09-513-996A-34109

Query Match      39.2%; Score 1037; DB 19; Length 514;
Best Local Similarity 44.1%; Pred. No. 2.8e-102;
Matches 209; Conservative 85; Mismatches 156; Indels 24; Gaps 10;

```

```

QY 6 ALPNDEWGPATAYOIEGAVKEGGRPSIWDTYCHLEBPTNGANGDYACDHYHRYDED 65
DB 29 SFRGAFVFGTASSAFQHEGAVKABRGFTIWDTSHTFGKITDPSNADVAVDQYHRYEED 88
QY 66 FDLITTKYKAYAPPSLSMSRIIPLGRLDPVNEGIEFYSKLIDALLRGITPWTLYHW 125
DB 89 VOLMKNNGMDAYRSISMTRIIPNG--VGHINEAGIDHYNKLINALAKGIEPYVTLYHW 146
QY 126 DLPOALHDIRYGMWLNVEEVLDFERYARLCEERGRDVRONMTINXPMIOAIYGYATGSN 185
DB 147 DLPOALHDIRYGMWLNPOIIN-DEFAVAEVCQFQRPGRVXKMTITFNEPHTALOGYDVLQ 205
QY 186 AGRSSINKIST--EGWTATEPMLAGKAQIMSHARAVAVYSRDFRPSQKQIGISLNGDY 243
DB 206 APGRCTIIFKLTCEGNSSTEPYIVGNVILTHATVSDIYRKKYKAKOGSLSGIAFDVMW 265
QY 244 YEPWDSNEPRDXEAAERMEFHIGMFANPIFLKDYEPESMKKOLGERLPALTPEADFAILN 303
DB 266 FEP-ESNKTEDI EAQRAQDFOLGMFLDPLMF-GDYPSMSRVSRLPVFTGSQSSLVK 323
QY 304 AGEUDFYGMNYTSGFARHLDPVPEPTDYLGAIHQENKDG-----SPVGEESGL 354
DB 324 -GSLDFVGINHYYTYARN-----NATNLIGTLHDAVSDSGTYTLPRKGLSTIGDRASS 377
QY 355 AMLRSCPDMPFRKHLARVYGLYK-PIYITENGCCPCGEENMTCEBAVNDPFRIRYFDSHL 413
DB 378 IWLIVPRGMRSLMNYIKHRYGNPVPFITENGMDPNSILSRKDALMDAKRIKYHHDYL 437
QY 414 DSISKAITODGVVYGYFAMALLDLEWSDGGRFGVFTDY--TTLKRTPKS 466
DB 438 SSLQASIKEDGCVKGYFVWSLIDNWEWAGISRFGLYFDYRDNLKRYPKDS 491

```

Search completed: March 25, 2004, 16:00:59  
 JOD time : 184 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 25, 2004, 15:55:42 ; Search time 15 Seconds  
(without alignments)  
540.737 Million cell updates/sec

Title: US-10-026-140-2

Perfect score: 2644  
Sequence: 1 MPESLALPNDPFWGFPATAY.....KSALVLKMPARQVAVAA 484

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 146766 seqs, 16758364 residues

Total number of hits satisfying chosen parameters: 146766

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PCR\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	34.9	571	6	US-10-767-701-46509
2	914.5	34.6	565	6	US-10-767-701-46537
3	865	32.7	567	6	US-10-152-372-254
4	857.5	32.4	1927	6	US-10-767-471-1083
5	826	31.2	438	6	US-10-093-037A-16
6	776	29.3	457	1	PCT-US04-02242-96
7	769	29.1	519	6	US-10-767-701-46260
8	765	28.9	462	1	PCT-US04-02242-94
9	702.5	26.6	478	6	US-10-417-884A-4922
10	633.5	24.0	415	6	US-10-767-701-36966
11	610.5	23.1	325	6	US-10-767-701-44803
12	607	23.0	495	6	US-10-417-884A-5715
13	589.5	22.3	470	6	US-10-417-884A-4517
14	567.5	21.5	378	6	US-10-767-701-37355
15	544.5	20.6	278	6	US-10-767-701-37578
16	540.5	20.4	244	6	US-10-767-701-39149
17	540.5	20.4	472	6	US-10-417-884A-6117
18	537	20.3	253	6	US-10-767-701-38415
19	502	19.0	420	1	PCT-US04-02242-90
20	502	19.0	454	6	US-10-093-037A-20
21	470.5	17.8	421	6	US-10-093-037A-17
22	457.5	17.3	485	6	US-10-417-884A-4797
23	449.5	17.0	480	7	US-60-551-121-10
24	446	16.9	503	6	US-10-417-884A-5013
25	445	16.8	437	1	PCT-US04-02242-104
26	437	16.5	210	6	US-10-767-701-39774

27	434.5	16.4	510	6	US-10-093-037A-22	Sequence 22, Appl
28	432	16.3	503	6	US-10-417-884A-6896	Sequence 6896, Ap
29	423.5	16.0	174	6	US-10-767-701-32266	Sequence 32266, A
30	418.5	15.8	497	6	US-10-417-884A-5279	Sequence 5279, Ap
31	414.5	15.7	509	6	US-10-093-037A-18	Sequence 18, Appl
32	410.5	15.5	511	6	US-10-093-037A-21	Sequence 21, Appl
33	409	15.5	484	1	PCT-US04-02242-68	Sequence 68, Appl
34	401.5	15.2	186	6	US-10-767-701-44337	Sequence 44337, A
35	400.5	15.1	512	6	US-10-093-037A-26	Sequence 26, Appl
36	393.5	14.9	512	1	PCT-US04-02242-114	Sequence 114, Appl
37	391	14.8	182	6	US-10-767-701-62456	Sequence 62456, A
38	376.5	14.2	476	1	PCT-US04-02242-20	Sequence 20, Appl
39	373.5	14.1	481	6	US-10-093-037A-15	Sequence 15, Appl
40	371.5	14.1	481	1	PCT-US04-02242-84	Sequence 84, Appl
41	366	13.8	162	6	US-10-767-701-62527	Sequence 62527, A
42	354	13.4	188	6	US-10-767-701-33603	Sequence 33603, A
43	347.5	13.1	314	6	US-10-767-701-43295	Sequence 43295, A
44	347	13.1	180	6	US-10-767-701-58259	Sequence 58259, A
45	343	13.0	171	6	US-10-767-701-62466	Sequence 62466, A

## ALIGNMENTS

RESULT 1  
US-10-767-701-46509  
Sequence 46509, Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53535)B  
CURRENT APPLICATION NUMBER: US/10/767, 701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 46509  
LENGTH: 571  
TYPE: PRT  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C54\_1.pep  
US-10-767-701-46509

Query Match	34.9%	Score 922;	DB 6;	Length 571;
Best Local Similarity	40.9%	Pred. No. 9.7e-78;		
Matches 200;	Conservative 87;	Mismatches 184;	Indels 18;	Gaps 13;
Qy	1	MPESLALPNDPFWGFPATAYQIEGAVKEGGGSIIMDTYCHLEPS-RTNGANGDVACDHY	59	
Db	70	IPKRDWFPSPFTTGAATSAFQIEGWNEDKSGSTMDHFCHTTPDFPIADSGNDVADSY	129	
Qy	60	HRVDEPDLTKYGAKAYRPSLSWSRIPLGRLDPVNEBIEFYSKLDALRGITPW	119	
Db	130	HLVDEPDLTKYGAKAYRPSLSWSRIPLGRLDPVNEBIEFYSKLDALRGITPW	188	
Qy	120	VTLVHMDLPALMDRYGKMLNVEVQDFERYARLCEREGDRVQVMTITNXPMIOAYG	179	
Db	189	VTLVHMDLPALMDRYGKMLNVEVQDFERYARLCEREGDRVQVMTITNXPMIOAYG	247	
Qy	180	YATGSNAPGSSINIGSTE--GMTATPEWLAGAOISHRRAVAVYSDRPSOKGOIGI	237	
Db	248	YGTGILAPGCSQGMKCPDPTGDSIRPEYLVGNHLLAHETVDLYNK-FHREKGIIGL	306	
Qy	238	SLNGDYVEPMDSEPRDKAERMEFHIGFANPIFLKKDYPSMKQGLERLPTPA	297	
Db	307	ALNMGVTPYGSFT-LDEQAHRCMDYNLGMYLEPV-VRGDYPHSMSSVDRLLPHPTK	364	
Qy	298	DFAILNAGEFDPFGMATYTSQFARHD----GPVETDVLGAIHEHQEKDQSPVEES	352	
Db	365	EQRKL-VGSDYDMIGINYSRFAKHVDITENFSPELNTDCCAT-EETTPNGNTIGPAT	422	

QY		355GLAMLRSCDPMRKHLARVYLGYK-PIYTENGCPCPGHEANNTCEBAVNDPRIRYPDS	411
Dd		423 GNAWLYLKGLDLIMMKRGRGNPNPYITTEGIDIDNGDISMRAALDDHRLDYLQR	482
QY		412 HUDSISKAITODGVVKKGFYAFALLDNLEMSDYGRRFGVTFTDYTT-LKRTPKSALVL	470
Dd		483 HI-SVLKDSDGANVRGHFTWSLLDNFEMSSSYTERFGIIVVDRENGCRTLKRSARML	541
QY		471 KMFAPAROR 479	
Dd		542 KEFNAGAKR 550	
 RESULT 2 US-10-767-701-46537			
; Sequence 46537, Application US/10767701			
; GENERAL INFORMATION:			
; APPLICANT: Kovacic, David K.			
; APPLICANT: Zhou, Yuhua			
; APPLICANT: Cao, Yongwei			
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With			
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement			
; FILE REFERENCE: 38-21(53535)B			
; CURRENT APPLICATION NUMBER: US/10/767, 701			
; CURRENT FILING DATE: 2004-01-29			
; NUMBER OF SEQ ID NOS: 63128			
; SEQ ID NO 46537			
; LENGTH: 565			
; TYPE: PRT			
; ORGANISM: Sorghum bicolor			
; FEATURE:			
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-C50_1.pep			
US-10-767-701-46537			
 Query Match            34.6%; Score 914.5; DB 6; Length 565;			
Best Local Similarity 40.7%; Pred.No.4,8e-77;			
Matches 198; Conservative 87; Mismatches 180; Indels 21; Gaps 13;			
QY		8 PNDFEWMGPATAAYQIEGAVKEGRGSPISMDTYCHLEPS-RINGANGDVACDHRYRDEDF	66
Dd		77 PPSFLFGAATSAHYQIEGAMNEDGKGPSWTMDHFCHNPPEMIIVDRSNGDAADSYHMVAELV	136
QY		67 DLITTKGAKARYFSLSMSRIITPLGGRLDPVNEEGIEFFSKLIDALLRIGITWVTLTYHD	126
Dd		137 RLKEMGWDAAFRSFISWPRIILPKGTLAGINKGEVEYNKLIDLLLENIGIEPYYITTFHMD	196
QY		127 LPQALHDHYGGWLAVEEVOLDPERFARLCFEERGGRVONWITINXPMWIAITYGATGSVA	186
Dd		197 TPQALVDAVYGGELD-ERIIKQDTYPAKYCFEKFGKKVKNWLFNFEBETFCASVGTGVLA	255
QY		187 PGRSS--INKSTSEGNTAITEPWLAGKAQIMSHARAVAAYSRDPRFSQKQIGISINDGY	244
Dd		256 PGRSGPGVSCAVPTENSLSSEPIYAHNLIRAHAEFTVIDYIN-K-YHGAGADRIGIALNVGR	314
QY		245 EPMDSNEPRDKEAAREMEFHGMFPANPLFKKOVPESKKQLGERRPALTPADRALINA	304
Dd		315 VPY-TNTLTLDQAQOARSMDKCIOWPLEPV-VRGDYPFSRARSARDURVPEYKEQEKL-V	371
QY		305 GETDFYGMNNYTSQFAHLD-----GPVETDYLGAIHEHOENK--DGSPVGEESGLAWL	357
Dd		372 GSYMIGGINYYTSPFSKHIDLSPNNSPLYLND---DAVASQETKPPDGNALGPPGNAMI	428
QY		358 RSCGDMFRKHLARVYGLYK-PIYTENGCPCPGHEANNTCEBAVNDPRIRIFYDSDLST	416
Dd		429 NMYPKGLHDLIMTMKNKYGNNPMYITTEGMGDIDKGDLPKPVALLDHRLLDIQRLSVL	488
QY		417 SKAITODGVVKKGFYAFALLDNLEMSDYGRRFGVTFTDYTT-LKRTPKSALVLKMF	475
Dd		489 KQSIDL-GADVGRGFANSLDNFEMSSSYTERFGIIVVDRENGCERTMKRSARMLQEFNG	547
QY		476 ARQRYK 481	
Dd		548 AKRVE 553	

```

RESULT 3
US-10-152-372-254
/ Sequence 254, Application US/10152372
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Beresini, Maureen
/ APPLICANT: DeForge, Laura
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerltisen, Mary E.
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Sherwood, Steven
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ TITLE OF INVENTION: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3330RIC394
/ CURRENT APPLICATION NUMBER: US/10/152,372
/ PRIOR FILING DATE: 2002-05-21
/ PRIOR APPLICATION NUMBER: 60/049911
/ PRIOR FILING DATE: 1997-06-18
/ PRIOR APPLICATION NUMBER: 60/056674
/ PRIOR FILING DATE: 1997-08-26
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059115
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059117
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059122
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059184
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059263
/ PRIOR FILING DATE: 1997-09-18
/ PRIOR APPLICATION NUMBER: 60/059352
/ PRIOR FILING DATE: 1997-09-19
/ PRIOR APPLICATION NUMBER: 60/059588
/ PRIOR FILING DATE: 1997-09-19
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 254
/ LENGTH: 567
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-152-372-254

Query Match      32.7%; Score 865; DB 6; Length 567;
Best Local Similarity 39.6%; Pred. No. 1,9e-72;
Matches 193; Conservative 73; Mismatches 179; Indels 42; Gaps 13;

QY      8 PNDFEMGATATAYOIEGAVKEGGRGPSIMDTYCHLEPRTNG-ANGDVACDHYRRYDEDF 66
Db      38 PLGFSWVGSSAYQTEGAMDDGDGSPSIWDVFTHSKGKQVLGNETADVACDGYTKVQEDI 97

QY      67 DLITRYGKAKAYRFSLSMSRIIPLCGRLDPVNEEGIEFYSKIDALLRRGITPWTLVYHD 126
Db      98 ILLELHNHNRFRFSLSWRLPFTGIRAEQVKKGIEFYSIDILALSSNIPITVLHHWD 157

QY      127 LPOLAHDPYGGWLVNEEYQUDFEERYARLCFRRPDGDRVQNTTIXXPMIQAIVGATGSA 186
Db      158 LPOLIQVYGGWNVNSMANT-FRDYANLCFAPFDGDRVQNTTISDPDPAAMEKGYETGHH 216

QY      187 PGRSINKGISTEGNTATEPWLAGKAQIMSHARAVALVVSBDPRPQKQIGISLNGDYEP 246

```

```

Db 217 PGLKLT-----RGTGLYKAAHHIIKAAKATWHSYNTTWRSKQOGLVSLNCWMBGP 267
Qy 247 WDSNEPRDKAARERBEFHIGFANPFLKKDYPESKKOLG-----ERLALTP 236
Db 268 VDISNPDLLEAERYLOFCIGWGFANPLY-AGDYPOVWKDYGKSAEQLEMSRLPVFSL 326
Qy 297 ADPAIINAGETDVGWNYYSOFARHLDDGVPPENDYLGAIEHOENKD-----GSPVGEES 352
Db 327 QEASYYK-CTSDPLGLGHFTTRITRYTERNYPSROGP-----SYONDRDLIELVDPMPDL 379
Qy 353 GLAMLRSCDPMFRKHLARVYLG-KRIYITENGCPGGEENMTCEAVNDPFRIRYFDS 411
Db 380 GSKMLVSPMGFRLINFAQTOYGDPRIYMEWG---ASQKHCTQ-LCDEMRIQYLGK 434
Qy 412 HLDISKAITODGVVVKGFAMALLDNLWESDGYGPRFGVTFPTYD--TLKRPKKSALV 469
Db 435 YINEMLKAI-KDGANIKGYSWSLDLDFEWKGYSDRYGYVEFNDRNKRPYKASVOY 493
Qy 470 LKDMFAA 476
Db 494 YKKIITA 500

```

## RESULT 4

```

US-10-767-471-1083
; Sequence 1083, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1083
; LENGTH: 1927
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-767-471-1083

```

Query Match 32.4%; Score 857.5; DB 6; Length 1927;

Best Local Similarity 40.3%; Pred. No. 4.9e-71;

Matches 192; Conservative 75; Mismatches 167; Indels 43; Gaps 12;

```

Qy 8 PNDPEWGATTAAYOIEGAVKEGGRGSPINDYCHLEPSRTNGANGVACDHYRYDEDFD 67
Db 1378 PEGFWSASAAVOIEGAWRADGKLSIMDTFSHTPLRVENDAIQDVACDSYHKIAEDLV 1437
Qy 68 LRTKYGAKAVRFLSRSRIIPLGRLDPVNEBGEIEFYSLIDALLRGITPWTLYHMDL 127
Db 1438 TLONLGVSHYRFSISMSRIIP-DGTRRYINEAGLNYVRLIDTLAASIOPOVITYHMDL 1496
Qy 128 POALHRYGGMWLNVEEVLDFERYARLCFERFGRVQNMWITINXPMIOAIYGYATGSNAP 187
Db 1497 PQLLOD-VGSMEN-ETIVQFKEXADVLFORLDKAKFMTILNEPFIATIQGIGYGAAP 1554
Qy 188 GRSSINKHSTEGNTATEPMLAGKAQIMSHARAVALVYSRDRFPQKQIGISLNGDYEPW 247
Db 1555 GVSN-----RPGTAPYIVGNHNLICAHAEAMHLVNDVYRASQGVISITISSDMAEPR 1606
Qy 248 DSNRPDKAERERMEFHIGFANPFLKKDYPESKKOLG-----FLPALTPA 297
Db 1607 DPNQDEVEAARYVQFMGWFAPHPFKNGDYNEVMKTRIRDSLAAGLKSRLPEFTES 1666
Qy 298 DPAIINAGETDVGWNYYSOFARHLDDGVPPETDYLGAIEHOENKDGSPVGE 353
Db 1667 EKRRIN-CTYDFGFGNHYTTVLAYNL-----NYATAISSFDADRGVAISADBSWPDG 1718
Qy 354 LAWLRSQPDMPFRKHLARVYLG-KRIYITENGCPGGEENMTCEAVNDPFRIRYFDSH 412
Db 1719 SPWLKMTVPFGFRRLINMLKEEYNDPPIYVTENGV-----SQREETDLNDTARIYLYRTY 1772

```

```

Qy 413 LDISKAITODGVVVKGFAMALLDNLWESDGYGPRFGVTFPTYD--TLKRPKKSAA 467
Db 1773 INEALKAV-QDKVLDLGRYTWMSAMDNFEMATGESERFGLHFVNYSDPSLPRIPKASA 1828

```

## RESULT 5

```

US-10-093-037A-16
; Sequence 16, Application US/10093037A
; GENERAL INFORMATION:
; APPLICANT: Jay M. Short
; APPLICANT: Bylina, Edward
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Mathur, Eric J.
; APPLICANT: Lam, David B.
; TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF
; FILE REFERENCE: 564462001402
; CURRENT APPLICATION NUMBER: US/10/093,037A
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: US 09/910,579
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 09/134,078
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 08/949,026
; PRIOR FILING DATE: 1997-10-10
; PRIOR APPLICATION NUMBER: US 60/056,916
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 438
; TYPE: PRF
; ORGANISM: Thermotoga sp.
US-10-093-037A-16

```

Query Match 31.2%; Score 826; DB 6; Length 438;

Best Local Similarity 38.4%; Pred. No. 5.8e-69;

Matches 183; Conservative 71; Mismatches 162; Indels 60; Gaps 14;

```

Qy 8 PNDPEWGATTAAYOIEGAVKEGGRGSPINDYCHLEPSRTNGANGVACDHYRYDEDFD 67
Db 8 PKDFIGTATTAAYOIEGAANEDRGPSIWDVSHPTGKTLNGDTGVACDHYRYXEDIQ 67
Qy 68 LRTKYGAKAVRFLSRSRIIPLGRLDPVNEBGEIEFYSLIDALLRGITPWTLYHMDL 127
Db 68 LMKELGLDLYRFSISMPRIIMPDKA--INQKVDFTNRLVDELKNDIIPFVTLYHMDL 124
Qy 128 POALHRYGGMWLNVEEVLDFERYARLCFERFGRVQNMWITINXPMIOAIYGYATGSNAP 187
Db 125 PVALYER-GSMNL-PIALYFRAVATFMNBLGDRVKWITILNEPWCSSFSGYTGEHAP 182
Qy 188 GRSSINKHSTEGNTATEPMLAGKAQIMSHARAVALVYSRDRFPQKQIGISLNGDYEPW 247
Db 183 GHQNLQ-----EAIITAAHNLREHGHAVQASREERK--DDEVGITVVMKIEPG 229
Qy 248 DSNRPDKAERERMEFHIGFANPFLKKDYPESKKOLGEBLPAITADPAIINAGET 307
Db 230 DA-KPESFLVASLVDKFNAMSHDPVFGK-YPEEVALYTEKGLOVLDSMNIIST-PI 286
Qy 308 DFGWNYYSOFA-----RHLDGVPETDYLGAIEHOENKDGSPVGEESGLAWL 357
Db 287 DFGWNYYSRTLVFPMNNPLGFSYVQGLPKTEWMEIY-----PQGLFDMLVYL 337
Qy 358 RSCPDMPFRKHLARVYLGKRIYITENGCPGGE-ENMTCEAVNDPFRIRYFDSHLS 416
Db 338 KE-----RYKLPLYITENGAGAPDKLEN-----GRVHDNYRIEYLEKFEKA 379
Qy 417 SKAITODGVVVKGFAMALLDNLWESDGYGPRFGVTFPTYD--TLKRPKKSALV 472
Db 380 LEAINAD-VDLKGYFWSLMDNFEMACGYSKRFGIIVYDYNTPKRIKDSAMWKE 434

```

## RESULT 6

```

PCT-US04-02242-96
; Sequence 96, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 96
; LENGTH: 457
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; NAME/KEY: DOMAIN
; LOCATION: (10) --(449)
; OTHER INFORMATION: Glycoeyl hydrolase family 1
PCT-US04-02242-96

```

```

Query Match      29.3%; Score 776; DB 1; Length 457;
Best Local Similarity 37.4%; Pred. No. 2.8e-64;
Matches 182; Conservative 67; Mismatches 171; Indels 66; Gaps 16;

```

```

QY 6 ALPNDPFWGFAATAYOIGAVKEGGGSPSMDTYCHLEPRTNGANGDVACDHYHRYDED 65
DB 12 SFPATFTGVATSAOIEGAALIGGRGSPIMDTFSHTEGKIIDGNSGVDACDHYHRYAED 71
QY 66 FDLTKYKAKYRFSLSMSRIIPGGRDLPVNEBEGIEFYSKLIIDALRGITPWTLYHW 125
DB 72 VELLASLDVNAVYRSMVSRVOPGSG--GANNEAGPDPYARLDLALAKGIDAHLLTYHW 129
QY 126 DLPQALHDYRGWLVNEEVQDLDFERYARLCFERGDRYQNMWITINXPMIOIYVATGNSN 185
DB 130 DLPQALQDE--GGWLN--RATCYHFAVAEVARRRFGHVASIATINNEPCITAVLGHGTGF 187
QY 186 ARGSSINXKSTEGNTATEPFLAKQAQ--IMSHARAVAVYSRDPRPSQKQIGISINDG 242
DB 188 APGAA-----DPAAVQVSHHLLSHGLMQAMRTYNABA--KLGIVLN-- 229
QY 243 YEEFMD-----SNEPRDKEAERMEFHIGWFANPIFLKDYEPESMKQIGERLPLATPAD 298
DB 230 ---QWTATPATDSQDRELALEIYARSVQYMDAIF--KGRYPALALKHIDRAQLSTFEND 285
QY 299 FALINAGETDFYGMNYYTSQPARHLDGVPETDYLAGIHEHQNKGDSPVGEESGLAMLR 358
DB 286 FIDIKQ--PIDLGVVYVYRFAVMSAETPRKRECKLGV-----NDMGW--E 327
QY 359 SCPDMFRKHARVYVGLV--KRIYITENGCC-----PCGGEENMTCEAVNDPFRIRYDSH 412
DB 328 TYPGGLTELLVGLHRETRLPVYITENGMAYADRPVNGK-----VHDEPREIYRHLH 379
QY 413 LDSISKAITODGVVVKGFAMALDNLMSDYGPRFGVFTDTYTLKRTPKKSATLVKD 472
DB 380 LDAL-RAVVAQGIYRGVYFWSLMDNFENMSGYAKRGMLVYDIATQCRSFKOSALMYRD 438
QY 473 MEAAQ 478
DB 439 FIAAQ 444

```

```

RESULT 7
US-10-767-701-46260
; Sequence 46260, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 46260
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-C4596_1.pep
US-10-767-701-46260

```

```

Query Match      29.1%; Score 769; DB 6; Length 519;
Best Local Similarity 38.3%; Pred. No. 1.5e-63;
Matches 179; Conservative 83; Mismatches 181; Indels 24; Gaps 14;

```

```

QY 8 PNDPFWGFAATAYOIGAVKEGGGSPSMDTYCHLEPRTNGANGDVACDHYHRYDED 67
DB 34 PAGVFGVGSAAVQVGAEDGRKPSIMDTFTHKYS--IDGATGVTDQYHKYKEDVK 92
QY 68 LITKYKAKYRFSLSMSRIIPGGRDLPVNEBEGIEFYSKLIIDALRGITPWTLYHWDL 127
DB 93 LHELGVDAYRMSIAMPRLIPDS--RGAVNPKGLEYYNNLIBELLSYGIQPHVTIYHPDF 150
QY 128 PQLAHDYRGWLVNEEVQDLDFERYARLCFERGDRYQNMWITINXPMIOIYVATGNSNAP 187
DB 151 PQLADEYKGLSPREI--DYTAIVADVCFKNGDRYKVSIVNEPIETIGFDDGILRP 209
QY 188 GRSS--INXSTEGNTATEPFLAKQAQIMSHARAVAVYSRDPRPSQKQIGISINDYYE 245
DB 210 QRSAPFGGCDGNSNTTBYIAHHLVLAHSAVSLYEKIQABEGGSIADALGMYE 269
QY 246 PWDNSPRDKEAERMEFHIGWFANPIFLKDYEPESMKQIGERLPLATPADFALINAG 305
DB 270 P-ATEPFDIIAAARNDPHVGFHMP--VGDYPPVVKVNGVSRLPSTNNE--AKRVKG 326
QY 306 EDPFYGMNYYTSQPAR---HLDGPVETDY--GAIHQ--ENKQSPVGEESGLAM 356
DB 327 SFDVGFNHYVAVYVADLSRLDEKV--RDYVADAVAADMPFLKRNQFPFGLGLTADF 384
QY 357 LNSCPDMFRKHARVYVGLVGP--IYITENGCCPGGEENMTCEAVNDPFRIRYPSHDS 415
DB 385 TITPFAALKMLKHQVYKINPAWIHENG--AAQDPDSGVNTYDDERSQYLDYIEA 442
QY 416 ISKAITODGVVVKGFAMALDNLMSDYGPRFGVFTDTYTLKRT 462
DB 443 TLQSI-RNGSNVQGYFVMSFLDVFEYLFGRYRFGVGVGFSTART 488

```

```

RESULT 8
PCT-US04-02242-94
; Sequence 94, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANI
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; CURRENT FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 94
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample

```

FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (9)...(462)  
OTHER INFORMATION: Glycosyl hydrolase family 1  
PCT-US04-02242-94

Query Match 28.9%; Score 765; DB 1; Length 462;  
Best Local Similarity 36.9%; Pred. No. 3e-63;  
Matches 176; Conservative 76; Mismatches 171; Indels 54; Gaps 16;

QY 5 LALPNDPEWGFATAOIEGAVKEGGRPSINDTYCHLEPSRTNGANGDVACHYHR 64  
DB LSPPEQFWGAAASVQVEGAVHEDGSPVMPCEKPGAVQGHGAAVACHYHRE 69  
QY 65 DFLLTGYGAKAYRFSLSWSRIIPLGRLDPVNEEGIEFYSKLIIDALLRGITPWTLYH 124  
DB DVALMQVGLHAYRLSVCPRVLPBG--VGQPNKGLDPSRLVDALLGKITPWTLYH 127  
QY 125 MDLPQALHRYGGLVNEVQDLFERVYARLCFERFGDRVQNMWTTINXPMIOAIYGYATGS 184  
DB MDYPLALYHR--GGMLNRDSADW--FAEYAGLIADRLSDRVQHFPTQNEPQYIGRHLEK 185  
QY 185 MARGSSINHGTEGNTATEPMLAGKAQIMSHAAVAVYSDRPPSGQIGISLNGDY 244  
DB HAEG-----DTLPMSCVLLAGHHSILAHGKAVALRA--QAKQQLRVGYAPVGMPL 234  
QY 245 EPWDSNEPRDKEAERMEFHI-----GWFANPIFLKDYPSMKQLGERLPALTP 236  
DB 235 HRP--TDSADVAAARKATFWREKSNMNMMDPFL--GETPAQGLAFGGDVQVRE 231  
QY 297 ADEAIIAGETDFYGMNYYTSQFARHLDPVPTDYLGAIHEQENKDGSPVEESGLAW 356  
DB 292 GDMQLI-AQPLDFGVNIVQSTPVR--ASSAESGF-----EVVPHPTGYPI--TAFNW 339  
QY 357 -----LRSCPDNFRKHLANVGLYKPIYITENGCCPCEENMTCEAVNDPRIRYFD 410  
DB 340 PITPOLYMGPRF-----YERYQPIYITENGLSC--RDVVAADGKVDARIDFTT 390  
QY 411 SHLDSISKATTOGVVVKGFAMALLDNLEMSDGYGRPGVTFTDYTLTKRTPKKA 467  
DB 391 RYLRHLRAVA-DGVAVEGTFHWSINDNFEMAGYRERFGLIHVDYETLAKTPKAS 446

RESULT 9  
US-10-417-884A-4922  
Sequence 4922, Application US/10417884A  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/417,884A  
FILING DATE: 17-Apr-2003  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4922:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 478 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (8) LOCATION 1...478  
SEQUENCE DESCRIPTION: SEQ ID NO: 4922:  
US-10-417-884A-4922

Query Match 26.6%; Score 702.5; DB 6; Length 478;  
Best Local Similarity 34.2%; Pred. No. 2e-57;  
Matches 169; Conservative 81; Mismatches 179; Indels 65; Gaps 15;

QY 3 ESALPNDPEWGFATAOIEGAVKEGGRPSINDTYCHLEPSRTNGANGDVACHYHR 62  
DB 5 ELTSPKDFLWGSASAAVQVEGAMQEDGSESWDFVRI PGKTFKGTGDLVDHYHR 64  
QY 63 DEFDLLTYGAKAYRFSLSWSRIIPLGRLDPVNEEGIEFYSKLIIDALLRGITPWTLYH 122  
DB 65 KEDIALMKQGLKAYRFSVAMTRIPNG--RGVNOAGLAFYERLDELLENIEEMTLT 122  
QY 123 YHMDLPQALHRYGGLVNEVQDLFERVYARLCFERFGDRVQNMWTTINXPMIOAIYGYAT 182  
DB 123 YHMDLPQALHRYGGLVNEVQDLFERVYARLCFERFGDRVQNMWTTINXPMIOAIYGYAT 181  
QY 183 GSNAPRSSINHGTEGNTATEPMLAGKAQIMSHAAVAVYSDRPPSGQIGISLNGD 242  
DB 182 AAHPG-----VTPDKRMYEVNHNLANANASVINKEHELEMPGKIGSPFAYT 228  
QY 243 YEPPDSNEPRDKEAERMEFHI--GWFANPIFLKDYPSMKQLGER--LPALTPADPA 300  
DB 229 PNYPIDSN--PENVALAENMEDLMANYKMD--VTMGKYPILAAKRFLEBKMAPITEAGDAE 286  
QY 301 ILNAGETDFYGMNYYTSQFARHLDPVPTDYLGAIHEQENKDGSPVEESGLAWLRSC 360  
DB 287 LLESARPDGLAINVY--QTATNAFNPL--NGVGAQKNTTGKGS--SEETGV----- 333  
QY 361 PMFRK-----HLAVYGLYKPIYITENGCCPCE--ENMTCE 397  
DB 334 PGMYKVENPFVERTNWEIDPEGLRIGLRITSRYPVLTENG--GEYDLTLPD 390  
QY 398 EAVNDPRIRYFDLSDISKATTOGVVVKGFAMALLDNLEMSDGYGRPGVTFTDYT 457  
DB 391 KOIHDDYRINTYLOSHKAIKEAIS--DGAVLDGCTMSYTDLSLNGYOKRIGFYVVDOD 449  
QY 458 ----TLKRTPKKS 466  
DB 450 ETQEGSLERYKKS 463

RESULT 10  
US-10-767-701-36968  
Sequence 36968, Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(53535)B



ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...495  
SEQUENCE DESCRIPTION: SEQ ID NO: 5715  
US-10-417-884A-5715

Query Match 23.0%; Score 607; DB 6; Length 495;  
Best Local Similarity 35.7%; Pred. No. 1,6e-48;  
Matches 177; Conservative 56; Mismatches 181; Indels 80; Gaps 24;

QY 7 LPNDFEWGATTAAYQIEGAVKEGGRPSIMDTYCHLEPSTNGANGDVACHYHRYDEDF 66  
DB 32 LPKDFIFGATTAAYQIEGATHTDQKGVANDKY--LEDNYWYTA--EPASDFHKKPVLD 87  
QY 67 DLTITGAKAYRSLMSRIIPLEGRLDPNBERGIERYSKIDALLRGITTPWTLTYHWD 126  
DB 88 ELAEKYGAVGIRISIMSRIFPTG--YGEVNEKGVETRYHNLFAECHKRHVEPVTLLHFD 145  
QY 127 LPQALHRYGGMIVNEVOLDFERYARLCPERFGDRVQNMITINXPMQAIYGATGNSA 186  
DB 146 TREVLIHSD--GDFLNRENIH--HFVDYAFCEEPPE--VNIWTTNEIGPIDGQTLVSKFP 202  
QY 187 PG-----RSSINGSTEGNTATEPWLGAQIMSHARAVALVSRDPRPSQKQIGI--SL 239  
DB 203 PGIQYDLAKVFGSHHN-----MMVSHARAVLYK---DKYKGEIGVYVHAL 245  
QY 240 NGDIYEPWDSNERDKEAERMEFHIGFANPIFLKKDYPSMK-----KOLGERLPA 293  
DB 246 PTXY--PYDPENPADYRAALEEDIINHKFLDATALYGHYSDKTMGEVNHILANGELD- 302  
QY 294 LTPADPAIINAGE--TDFYGMNYYTSQFARHLDPVPETDYLGAIHEHONKNDGSPVGE 351  
DB 303 LRPEDQALERAQDLNDFLGINYYMSDWMQAPDG--ETG--IINHGKEK--GSSKTYQI 355  
QY 352 SGLAWLRSCPEMRKILAR-----VY--GLYG-----KPIYITENGCPGPE 391  
DB 356 KGVG--RIAPD-----YVPRTDWMIITPEGLYQIMKVNKDYPNKKIYITENGL---GY 407  
QY 392 ENMTCEAVNDPRIRYFDSHLDSISKAITQDGVVVKGYAMALLDNLEMSDGYGPRFGV 451  
DB 408 KGEFVNTVYDDDRIRIYVKKHLEVLSPALA--DGANVKGYPFIMSLMDFVSMNGEKRYGL 466  
QY 452 TETDYTTLKRTPKKSA 467  
DB 467 FYVDFDTQERYPKSA 482

RESULT 13  
US-10-417-884A-4517  
Sequence 4517, Application US/10417884A  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSES: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/417, 884A  
FILING DATE: 17-Apr-2003  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: CTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4517:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 470 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (B) LOCATION 1...470  
SEQUENCE DESCRIPTION: SEQ ID NO: 4517;  
US-10-417-884A-4517

Query Match 22.3%; Score 589.5; DB 6; Length 470;  
Best Local Similarity 32.9%; Pred. No. 6,4e-47;  
Matches 168; Conservative 75; Mismatches 161; Indels 107; Gaps 25;

QY 5 LALPNFENGATTAAYQIEGAVKEGGRPSIMDTYCHLEPSTNGANGDVACHYHRYDE 64  
DB 6 LKLPEDIFGATTAAYQIEGATHTDQKGVANDKY--LEDNYWYTA--EPASDFHKKPVLD 87  
QY 65 DPLITGAKAYRSLMSRIIPLEGRLDPNBERGIERYSKIDALLRGITTPWTLTY 123  
DB 62 DIELGERFVNGIRISIMSRIFPDGAKRNP--EGIAFYHRVFECKRANTPVTLLH 118  
QY 124 HMDLPQALHRYGGMIVNEVOLDFERYARLCPERFGDRVQNMITINXPMQAIYGATG 183  
DB 119 HFDTPRLPD--HGDPLNRETIH--FVSVAIFCFHEPKE--VKWSTNEIETIPVATNOYLLG 175  
QY 184 SNARG--RSSINK-----HSTEGNTATEPWLGAQIMSHARAVALVSRDPRPSQKQIGI 237  
DB 176 VPPGIGIKYDFTKIYVACLHN-----MMVAHARVNVYKNEHP--GEIGV 217  
QY 238 --SLANGDYEPWDSNERDKEA--ERRMEF--HIGMFANPIFLKKD----- 278  
DB 218 VHSLETKYAA--TDAPEBKHAFLDDALISIRFLDATALYGHYSTETLTLALDEICEANQA 274  
QY 279 ---YPSMKKQIGERLPALTLPADPAIINAGE--DFYGMNYYTSQFARHLDPVPETDY 333  
DB 275 SYHFPPE-----DFVELKKASTDNDYGINHYOCHVKAADGE----- 312  
QY 334 GAHHEQNKDGSPPVEESGLA-----WLSCPMPFKHLARV---GYLG 376  
DB 313 NAIHNHGTEKGSYVKKVIGIRIYKEGIPTDMDL--IYEGLYDLLRLRISDPHYN 371  
QY 377 KPIYITENGCPGGEENMTCEAVNDPRIRYFDSHLDSISKAITQDGVVVKGYFAMALL 436  
DB 372 K-IYITENGM---GYDQFEDGIMQPRIDYLRVYVLESLSKAITA--GVNVKGYFLMSLM 426  
QY 437 DNLEMSDGYGPRFGVTFDTYTTLKRTPKKSA 467  
DB 427 DLFSTMTNGYNKRYGLFYVDFETQKRYPKESA 457

RESULT 14  
US-10-767-701-37355  
Sequence 37355, Application US/10767701  
GENERAL INFORMATION:

APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 37355  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-C2562\_1.pep  
US-10-767-701-37355

Query Match 21.5%; Score 567.5; DB 6; Length 378;  
Best Local Similarity 35.5%; Pred. No. 5.4e-45;  
Matches 125; Conservative 66; Mismatches 150; Indels 11; Gaps 10;

QY 118 PWTLYHMDLPQALHRYGGMILNVEVOLDPERFARLCFEERFGDRVQNMWITINXPMWIOAI 177  
DB 4 PVTINHYDIPBELQOORYGSMWS-PEIQEDFTYFAELCFKMFGRVXKMAITENPNIIMAK 62  
QY 178 YGATGSNAPGRSS--INKHSTEGNTATEPWLAKQAQIMSHARAVALVYSRDPRPSQKQOI 235  
DB 63 IAYFGKPPSPHCSKPFGR-CNSGNSSTEPYLAHNMILAAKTVNIYKKVYKTKOGGSV 121  
QY 236 GISLNGDYEPWDSNEPRDKAEARRMEFHIGFANPIFLKDYVESMKKOLGERPLALT 295  
DB 122 GITVWRWYEP- SVYTDHILAVSRAQSFAPWFLDPLFF-GDYPHQMRQILGPNLPFT 179  
QY 296 PADPAILNAGDEDFVGMNYTQSOPARHLDGPVETD-YLG-AIHENKDKGSPVGEESG 353  
DB 180 EGEKQLMK-NQIDFISVNHKTLVDCVYSLCDDLTVDGALVSESAERNIGIPGKPTP 238  
QY 354 IAWLRSCPDMPFRGHARVYGLYK-PIYITENGCCPCGEEENMTCEAVNDPRIRYPSH 412  
DB 239 VANNVVPSSMELVWYLTQRYKSIPLYITENGVAQIGNSPPTBELINDTERSSYINDY 298  
QY 413 LDSISALITODGVVYKGFYFAMALDNLLEWSDGGRFGVTFTDYTLIKRTPK 464  
DB 299 LTVLSFAL-RKGADVGYFWMSLMDNFEWLSGITTKYGFYHDFKSLKRTPK 349

## RESULT 15

US-10-767-701-37578  
Sequence 37578, Application US/10767701

GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701

CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128

SEQ ID NO 37578  
LENGTH: 278

TYPE: PRT  
ORGANISM: Sorghum bicolor

FEATURE:  
NAME/KEY: unsure

LOCATION: (1) .. (278)  
OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: SORBI-28MAY03-C31549\_1.pep  
US-10-767-701-37578

Query Match 20.6%; Score 544.5; DB 6; Length 278;  
Best Local Similarity 44.7%; Pred. No. 4.9e-43;

Matches 109; Conservative 45; Mismatches 79; Indels 11; Gaps 5;  
QY 24 GAVEGGRGPSIMDTYCHLEPSPRTNGANGVACDHYHRVDEBDLITKYGAKAYRFSLSW 83  
DB 46 GAREGGKGSIMDVFTDDKERVLDKSMAEIAVDHYRKYEDIELMASLGFSAIRFSISW 105  
QY 84 SRITPLGGRLDPVNEEGIEFYSKLIDALRRGITPWLYHMDLPQALHRYGGMILNVEE 143  
DB 106 ARIFP-DLGEKVEQVAFINDLINMISKGIEPVATLYHMDLPNNLQKTLGGWISDKI 164  
QY 144 VQDPERFARLCFEERFGDRVQNMWITINXPMWIOAIYGATGSNAPGRSSINKHSTEGNTAT 203  
DB 165 VEY-FALVAACFANFGDRVGRWITINEPLOTALNGIGIFAFG-----GCQGETA- 215  
QY 204 EFWLAGKAQIMSHARAVALVYSRDPRPSQKQOIGISLNGDYEPWDSNEPRDKAEARRME 263  
DB 216 RCTLAHHQILAAHAAVAVYRRKFAAGGAGVGVDCENAFEP-SEKAEQIAAGRRID 274

Search completed: March 25, 2004, 16:01:26  
Job time : 16 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:57:07 ; Search time 5341.5 Seconds  
(without alignments)  
11130.890 Million cell updates/sec

Title: US-10-026-140-1  
Perfect score: 1991  
Sequence: 1 agcagcgtccgacacagcag.....ttttaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_pmg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	702.2	35.3	773	14	CB904679 trico39xe
2	701	35.2	703	14	CF875727 trico39xe
3	461.2	23.2	584	14	CF884351 trico39xe
4	387	19.4	1994	9	AA415086 MG0020 RC

5	201	10.1	2031	11	AY109400
6	195.2	9.8	788	14	CB677158 OSUNDEL3N
7	193	9.7	786	14	CB673495 OSUNGE08A
8	190.4	9.6	806	14	CB642881 OSUNEB03E
9	179.2	9.0	889	14	CD438244 EL01N0511
10	175.6	8.8	480	14	CB688082 CBST-23-E
11	175	8.8	1182	11	AY106991 Zea mays
12	171.8	8.6	662	14	CF880605 trico82xh
13	171.8	8.6	732	14	CB907493 trico82xh
14	169.8	8.5	650	10	BE77259 MY-26-C-1
15	162.6	8.2	613	9	AJ436490 AJ436490
16	158.6	8.0	860	14	CF666831 RTCNT1_26
17	157.8	7.9	610	13	BO1105D5 BO1105D5
18	157	7.9	825	14	CK200642 FGAS00915
19	156.2	7.8	1126	14	CK209455 FGAS02122
20	155.8	7.8	708	13	CA145286 SCSGRT206
21	155.6	7.8	673	14	CD482402 atc01-26m
22	155.6	7.8	674	14	CD458246 FG08_06h0
23	154	7.7	786	14	CB634193 OSIEB131
24	153.8	7.7	1000	14	CF933596 TRESB-B08
25	153.4	7.7	732	14	CB633739 OSIEB12M
26	153.4	7.7	808	14	CB633785 OSIEB12M
27	153.2	7.7	739	13	CA139881 SCEZRT201
28	153	7.7	1126	14	CK161716 FGAS01428
29	152.4	7.7	788	14	CF554428 G9M12 Inf
30	151.8	7.6	728	14	CF473575 RTW2_3 H
31	151.6	7.6	723	12	BM605139 170006870
32	151.6	7.6	766	14	CB643857 OSUNEB04M
33	151.6	7.6	821	14	CB634605 OSIEB14D
34	151.2	7.6	564	14	CB641046 OSUNEB17J
35	151.2	7.6	636	9	AV913799 AV913799
36	151	7.6	676	12	BM621029 170006874
37	151	7.6	805	14	CB677500 OSUNEB14I
38	150.8	7.6	646	12	BJ235347 BJ235347
39	150.8	7.6	839	14	CK201813 FGAS01033
40	150.6	7.5	829	14	CB663657 OSUNED080
41	150	7.5	592	14	CA193361 SCRLAD104
42	150	7.5	760	14	CB643189 OSUNEB03L
43	150	7.5	794	14	CB654045 OSUNEB05K
44	150	7.5	806	14	CB674932 OSUNEB10E
45	150	7.5	826	14	CB648777 OSUNEB12P

## ALIGNMENTS

RESULT 1  
CB904679 773 bp mRNA linear EST 02-JUL-2003  
LOCUS  
DEFINITION  
trico39xe13 T. reesei mycelial culture, Version 3 apr11 Hypocrea  
jecorina cDNA clone trico39xe13, mRNA sequence.

ACCESSION  
CB904679 GI:30119337

VERSION  
CB904679.1

KEYWORDS  
EST.

SOURCE  
Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM  
Hypocrea jecorina

REFERENCE  
Bukaryova; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaeae; Hypocrea.

AUTHORS  
1 (bases 1 to 773)  
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., Englund, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.

TITLE  
Transcriptional regulation of biomass-degrading enzymes in the  
filamentous fungus Trichoderma reesei

JOURNAL  
J. Biol. Chem. 278 (34), 31988-31997 (2003)

MEDLINE  
22803314

PUBMED  
12788920

COMMENT  
Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817

Email: Pforeman@genencor.com  
Seq primer: LR-F1 primer.  
Location/Qualifiers  
FEATURES  
source

1. .773  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="Qm6a"  
/db\_xref="taxon:51453"  
/clone="tric039x13"  
/dev\_stage="mycelia"  
/clone\_lib="T. reesei mycelial culture, Version 3 april"  
/note="Vector: pREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 35.3%; Score 702.2; DB 14; Length 773;  
Best Local Similarity 92.8%; Pred. No. 4.8e-114;  
Matches 707; Conservative 1; Mismatches 54; Indels 0; Gaps 0;

340 GGATCCCGTCAACGAGGAGGAAATTGAGTTTAAACGAACTGATTGACGCTTGTGAG 399  
12 GNNNCTGCTGCTGTAANNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 71  
400 GCGGGGATACAGCTTGGGTGACTTTGTACCACTGGGATCTGCTCGAGGGCTTCAAGA 459  
72 GCGGGGATACAGCTTGGGTGACTTTGTACCACTGGGATCTGCTCGAGGGCTTCAAGA 131  
460 TCGGATGAGGCTGCTCAACGAGGAGGTCAGCTGAGCTTTGACGGGTATGCGAG 519  
132 TCGGATGAGGCTGCTCAACGAGGAGGTCAGCTGAGCTTTGACGGGTATGCGAG 191  
520 GTTGTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAAGMCCCTG 579  
192 GTTGTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAAGMCCCTG 251  
580 GATTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639  
252 GATTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311  
640 CAAGCACTCCACGAGGAGGCAACCTGCACTGAGCGCTGCTGCGAAAGGCCAAGT 699  
312 CAAGCACTCCACGAGGAGGCAACCTGCACTGAGCGCTGCTGCGAAAGGCCAAGT 371  
700 CATGAGCCATGCCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 759  
372 CATGAGCCATGCCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 431  
760 CGAGTCGGCATCTCGCTCAACGCGGACTTATGAGCCCTGGGACAGCAATGAGCTTG 819  
432 CGAGTCGGCATCTCGCTCAACGCGGACTTATGAGCCCTGGGACAGCAATGAGCTTG 491  
820 GGAAGAAGAGGCTGCTGAGCGAGATGGAATTCATTGGCTGGTTGGCAATCCCAT 879  
492 GGAAGAAGAGGCTGCTGAGCGAGATGGAATTCATTGGCTGGTTGGCAATCCCAT 551  
880 CTCTTGAAGAAGACTATCAAGAGCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939  
552 CTCTTGAAGAAGACTATCAAGAGCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 611  
940 CTCTACTCCGCGGAGCTTTGCAATCTCAATGCGGAGAGAGAGAGAGAGAGAGAGAG 999  
612 CTCTACTCCGCGGAGCTTTGCAATCTCAATGCGGAGAGAGAGAGAGAGAGAGAGAG 671  
1000 TTAATAACATCCAGTTTCCGCGGCACTTGAAGAGTCCGCTCCGAGAGCGAGATATCT 1059  
672 TTAATAACATCCAGTTTCCGCGGCACTTGAAGAGTCCGCTCCGAGAGCGAGATATCT 731  
1060 CGGCGCATCATGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1101  
732 CGGCGCATCATGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 773

RESULT 2  
CF875727 703 bp mRNA linear EST 31-OCT-2003  
LOCUS  
DEFINITION  
tric039x13.b1 T. reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric039x13, mRNA sequence.  
CF875727  
CF875727.1 GI:38130409

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Hypocrea jecorina (anamorph: Trichoderma reesei)  
Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE  
AUTHORS  
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,  
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and  
Dean, R.A.

TITLE  
JOURNAL  
COMMENT  
Analysis of the protein processing and secretion pathways in a  
Trichoderma reesei EST dataset  
Unpublished (2003)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu

Seq primer: LR-F1 primer.  
Location/Qualifiers

FEATURES

source

1. .703  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="Qm6a"  
/db\_xref="taxon:51453"  
/clone="tric039x13"  
/dev\_stage="mycelia"  
/clone\_lib="T. reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3Y; Site\_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 35.2%; Score 701; DB 14; Length 703;  
Best Local Similarity 99.7%; Pred. No. 8.1e-114;  
Matches 701; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

399 GCGGGGATACAGCTTGGGTGACTTTGTACCACTGGGATCTGCTCAAGCGCTTACG 458  
1 GCGGGGATACAGCTTGGGTGACTTTGTACCACTGGGATCTGCTCAAGCGCTTACG 60  
459 ATCGCTATGAGGCTGCTCAACGAGGAGGTCAGCTGGAATCCATCAACGAMCCTT 518  
61 ATCGCTATGAGGCTGCTCAACGAGGAGGTCAGCTGGAATCCATCAACGAMCCTT 120  
519 GATTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAACGAMCCTT 578  
121 GATTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAACGAMCCTT 180  
579 GATTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAACGAMCCTT 638  
181 GATTGCTTGAAGCTTTTGGGACCGAGTCCAGAACTGGATCACCATCAACGAMCCTT 240  
639 ACAAAGCTTCCACGAGGAGGCAACCTGCACTGAGCGCTGCTGCTGAGAAAGGCCAGA 698  
241 ACAAAGCTTCCACGAGGAGGCAACCTGCACTGAGCGCTGCTGCTGAGAAAGGCCAGA 300  
699 TCATGAGCATGCCGCGCGCTGGCGCTTCAAGAGGAGACTTTGCGCCCTTGGCAAAAG 758  
301 TCATGAGCATGCCGCGCGCTGGCGCTTCAAGAGGAGACTTTGCGCCCTTGGCAAAAG 360  
759 GCCAATGCGCATCTGCTCAACGCGGACTTATGAGAGCCCTGGGAGAGAGAGAGAGAG 818  
361 GCCAATGCGCATCTGCTCAACGCGGACTTATGAGAGCCCTGGGAGAGAGAGAGAGAG 420

QY 819 GGGACAAAGAGGCTGCTGAGCGGACGATGGATTTCACATTGGCTGGTTTGCCATGCCA 878  
 DB 421 GGGACAAAGAGGCTGCTGAGCGGACGATGGATTTCACATTGGCTGGTTTGCCATGCCA 480  
 QY 879 TCTTTCTTGAAGAAGACTATCCAGAGAGATGAAGAAGAGCTGGCGAGAGGCTTCCAG 938  
 DB 481 TCTTTCTTGAAGAAGACTATCCAGAGAGATGAAGAAGAGCTGGCGAGAGGCTTCCAG 540  
 QY 939 CCCTCACTCCCGGAGCTTTGGCATCTCTCAATCCCGAGAGACCGACTTCTACGCGATGA 998  
 DB 541 CCCTCACTCCCGGAGCTTTGGCATCTCTCAATCCCGAGAGACCGACTTCTACGCGATGA 600  
 QY 999 ATTACACACATCCCGAGCTTGGCGGCGACCTAGACGGTCCCGCCCGAGAGGAGCATATC 1058  
 DB 601 ATTACACACATCCCGAGCTTGGCGGCGACCTAGACGGTCCCGCCCGAGAGGAGCATATC 660  
 QY 1059 TCGGCGCCATCCATGAGCACCAGAGAAATMAGACGGCAGCCC 1101  
 DB 661 TCGGCGCCATCCATGAGCACCAGAGAAATMAGACGGCAGCCC 703  
 RESULT 3  
 CF884351 584 bp mRNA linear EST 31-OCT-2003  
 LOCUS tric039xel3.b12 T. reesei mycelial culture, Version 6 October 2003  
 DEFINITION Hypocrea jecorina cDNA clone tric039xel3, mRNA sequence.  
 ACCESSION CF884351  
 VERSION CF884351.1 GI:38139033  
 KEYWORDS EST.  
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
 ORGANISM Hypocrea jecorina  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
 1 (bases 1 to 584)  
 Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houlek, T.D.,  
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and  
 Dean, R.A.  
 Analysis of the protein processing and secretion pathways in a  
 Trichoderma reesei EST dataset  
 Trichoderma reesei  
 Unpublished (2003)  
 Contact: Ralph A. Dean  
 Fungal Genomics Laboratory  
 North Carolina State University  
 Campus Box 7251, Raleigh, NC 27695, USA  
 Tel: 919-513-0020  
 Fax: 919-513-0024  
 Email: ralph.dean@ncsu.edu  
 Seg primer: Tr-F1 primer  
 Location/Qualifiers  
 1..584  
 /organism="Hypocrea jecorina"  
 /mol\_type="mRNA"  
 /strain="QMGa"  
 /db\_xref="taxon:51453"  
 /clone="tric039xel3"  
 /dev\_stage="mycelia"  
 /clone\_1lb="T. reesei mycelial culture, Version 6 October  
 2003"  
 /note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial  
 culture grown from 24 hrs to 6 days with varying Carbon  
 and Nitrogen sources and concentrations."

## ORIGIN

Query Match 23.2%; Score 461.2; DB 14; Length 584;  
 Best local Similarity 94.0%; Pred. No. 2e-71;  
 Matches 489; Conservative 1; Mismatches 29; Indels 1; Gaps 1;

QY 399 GGGGGGATACAGCCTTGGGTGATTTGACACATGGGATCTGCTCAGGCGCTTACAG 458  
 DB 1 GGGGGGATACAGCCTTGGGTGATTTGACACATGGGATCTGCTCAGGCGCTTACAG 60  
 QY 459 ATGGCTATGAGAGCTGGCTCAAGCTGGAAGAGTCCAGCTTGACCTTGACGCGATGCGA 518

DB 61 ATGGCTATGAGAGCTGGCTCAAGCTGGAAGAGTCCAGCTGAGCTTTGAGCGGTATGCGA 120  
 QY 519 GGTGTGCTTTGAACGTTTGGGAGACCGAGTCCAGAACTGATCACCATCAACGAAACCT 578  
 DB 121 GGTGTGCTTTGAACGTTTGGGAGACCGAGTCCAGAACTGATCACCATCAACGAAACCT 180  
 QY 579 GGATTGAGGCACTATATGATATATGACCGGAGCAACGCCCGGCGAGAGCAGATT 638  
 DB 181 GCATTGAGGCACTATATGATATATGACCGGAGCAACGCCCGGCGAGAGCAGATT 240  
 QY 639 ACAAGCACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGCTGGAAGGCCAGA 698  
 DB 241 ACAAGCACTCCACCGAGGCAACACTGCACTGAGCCGTGCTCGCTGGAAGGCCAGA 300  
 QY 699 TCATGAGCCATGCGCGCGCGCGCTGACGACGAGGACTTCCGCTCCGCAAAAG 758  
 DB 301 TCATGAGCCATGCGCGCGCGCGCTGACGACGAGGACTTCCGCTCCGCAACGG 360  
 QY 759 GCCAGATCGGCATCTGCTCAACGCGGACACTATGAGCCCTGGAGACGATGAGCTC 818  
 DB 361 GCCAGATCGGCATCTGCTCAACGCGGACACTATGAGCCCTGGAGACGATGAGCTC 420  
 QY 819 GGGACAAAGAGGCTGCTGAGCGGACGATGGAAATTTGACATTGGCTGTTG-CCAAATCCC 877  
 DB 421 GGGACAAAGAGGCTGCTGAGCGGACGATGGAAATTTGACATTGGCTGTTG-CCAAATCCC 480  
 QY 878 ATCTTTCTTGAAGAAGACTATCCAGAGAGCATGGAAGAGC 917  
 DB 481 ATCTTTCTTGAAGAAGACTATCCAGAGAGCATGGAAGAGC 520

RESULT 4  
 AA415086 1994 bp mRNA linear EST 09-DEC-1999  
 LOCUS Mg0020 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone  
 DEFINITION RCW20 similar to Beta-Glucosidase (EC 3.2.1.21), mRNA sequence.  
 ACCESSION AA415086  
 VERSION AA415086.1 GI:2537251  
 KEYWORDS EST.  
 SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
 ORGANISM Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes; Incertae sedis; Magnaportheaceae; Magnaporthe.  
 1 (bases 1 to 1994)  
 Wu, S.-C., Bernstein, B.D., Dervill, A.G. and Albersheim, P.  
 Expressed sequence tags of the rice blast fungus grown on rice cell  
 walls  
 Unpublished (1997)  
 Contact: Sheng-Cheng Wu  
 CCRC  
 University of Georgia  
 220 Riverbend Road, Athens, GA 30602-4712, USA  
 Tel: 706 542 4446  
 Fax: 706 542 4412  
 Email: wus@ccrc.uga.edu  
 Fully sequenced  
 Insert Length: 1994 Std Error: 0.00.  
 Location/Qualifiers  
 1..1994  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /strain="CP987"  
 /db\_xref="taxon:148305"  
 /clone="RCW20"  
 /sex="Mati-1 hermaphrodite"  
 /tissue\_type="Mycelium"  
 /dev\_stage="Day 5 post-inoculation"  
 /clone\_1lb="RCW Lambda Zap Express Library"  
 /note="Vector: pBluescript excised from Lambda Zap  
 Express; Site 1: EcoRI; Site 2: XhoI; Day 5  
 post-inoculation mRNAs prepared from Magnaporthe grisea  
 grown at 23C in the dark with constant gyratory shaking

## FEATURES

source  
 1..1994  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /strain="CP987"  
 /db\_xref="taxon:148305"  
 /clone="RCW20"  
 /sex="Mati-1 hermaphrodite"  
 /tissue\_type="Mycelium"  
 /dev\_stage="Day 5 post-inoculation"  
 /clone\_1lb="RCW Lambda Zap Express Library"  
 /note="Vector: pBluescript excised from Lambda Zap  
 Express; Site 1: EcoRI; Site 2: XhoI; Day 5  
 post-inoculation mRNAs prepared from Magnaporthe grisea  
 grown at 23C in the dark with constant gyratory shaking

100 rpm in Vogel's minimal medium containing 0.5% isolated rice cell walls as the sole carbon source. Library provided by Sheng-Cheng Wu. Sequences were processed by one of two methods. Where a full-length alignment to the M. grisea genome sequence was available, the EST sequence was trimmed according to the alignment, otherwise sequence quality was assessed using phred/Phrap version 991019 and trimmed according to phd files (0.05) and for vector seqs."

## ORIGIN

Query Match 19.4%; Score 387; DB 9; Length 1994;  
Best Local Similarity 57.0%; Pred. No. 1.8e-58;  
Matches 806; Conservative 1; Mismatches 591; Indels 15; Gaps 5;

```

Qy 78 CTCGCCCAACGACTTTGAAATGGGGCTTCGACAGCGCCGCTTACCAATCGAAGCCCG 137
Db 290 CTCTCCCAAGAGATTTCTTTGGGGCTTCGACAGCGCTCATACAGATCGAGGGTCTA 349
Qy 138 TCAAGAGAGTGGCCGCGCCGCTTCATCTGGACAGCTACTGCACCTGGAGCCATCGC 197
Db 350 TCGACAGAGATGTCGTGGGCCGCTCCATTGGGATACCTTACTGCATTTCCGGGTAAAG 409
Qy 198 GCACCAACGCGCCCAACGCGCATGTGGCTTCGATCACTACACCGCTTACGATAGAGCT 257
Db 410 TCGCAGACGCGACGCTCGGGTGTGACGGGATCGACTCATCAACAGAGACGAGAGACA 469
Qy 258 TTGATCTCTTGACCAAGTAAGCGCCAAAGGCTTACCGCTTCTTGTGCGGTGCGGGA 317
Db 470 TTGACCTGCTCAAGTCTGTTGGTCCCAAGTGTATCGCTTTTCATCTCGTGTGAGGA 529
Qy 318 TCATTCCTCTCGCGCGAGGCTGGATCCGTCACAGAGAGGAATGATTTTACAGCA 377
Db 530 TTATCCCATATGGTGGCCGTATATGACCCCATCAACAGAGAGGATCGAACCATACGTCA 589
Qy 378 AACTGATTTGACGCTCTGTTGAGCGGGGTATCAGCCTTGGGTACTTTGTACACTGGG 437
Db 590 AGTTCGTAGACGACTGCTCGAGGCGCGCATCAGCCATCATACCTCTTCCACTGGG 649
Qy 438 ATTCGCTCAGGGGCTTCACAGATGCTATGAGAGCTGGCTCAAGTGAAGAGGTCCAGC 497
Db 650 ATTCGCGGAGATGATTGACAGAGGATGAGGTGGTGTCTGTAACAGAGAAAGTTCCCT 709
Qy 498 TGAATTTGAGCGGTATGCGAGGTTGTCTTGAACGTTTGGGAGCAGATCCAGAACT 557
Db 710 TGAATTTGAGCACTATGACAGGCTCATGTTCAGAGGCATTTCCAAAGT---CAAACCT 766
Qy 558 GGATCACCATTAAAGAACCTTGATTCAGGCCATCTATGATATGCAACGGCAGCAACG 617
Db 767 GGATCACCCTTCMAAGAGCGGTGTGCTGTCTCATTTCTCGCTTACAGCGTGGTGAAGTTG 826
Qy 618 CCGCGGAGAGAGAGCATTAACAGACATCCACGAGGGCAACATGCCATCGACAGCGT 677
Db 827 CGCTGGCGCTGTCTCGACCGCAGCAAGTGTCTGTGCGGAGACGCTCTCGGAGGCT 886
Qy 678 GGCCTGCTGAAAGCCCAATCATAGCCATGCGCGCGCGTGGCGCTTACAGCAGAGG 737
Db 887 GGATTTGGGGCCCAACCTTCTGTGGGCCACGTCGTGCGCTCAAGGTGACAGGAG 946
Qy 738 ACTTTGCGCTTCGCAAAAGGGCCAGATCGGCATCTGCTCAACGGGACTACTATAGAC 797
Db 947 AATTCAAGGCGCAGACAGAGGTGATGTTGATACAGTTGAATGGGACCCCACTTCC 1006
Qy 798 CCGTGGACAGCAATGAGCTCGGACAAAGAGGTGCTGAGCGACGAGTGAATTTTACA 857
Db 1007 CTTGGATTCGGAAGACCCCAAGGAGCGTGTATGACACCAACGCAAGTTCAGATTCCGCA 1066
Qy 858 TTGGCTGTTTGCATTCCTCTTGAAGAGAGACTATCCAGAGAGCATGAAGAGC 917
Db 1067 TCTGTGTTTGGGACCCCATCTACTTTGGCGAG---TATCCAGTCTCGATGTGTAAGC 1123
Qy 918 AGCTGGGAGAGGCTTCAGCCTCACTCCCGGAGCTTTGCAATCTCAATCCGAG 977

```

```

Db 1124 AGCTGGCGACCGCTGCCCCACTTTTACAGAGAGAGAGAAAGCTTTAGTCAA---GGGTT 1180
Qy 978 AGACCGACTTCTACGGCATGAATTTACTACACATCCCGATTCGGCGGCACCTAGACGGTC 1037
Db 1181 CAACGACCTTTATGATATACACTGCTACCGGCACTACATCAGGACCAAGAGAGAG 1240
Qy 1038 CCGTCCCGAGAGAGCACTATCTCGGCGCATTCATAGCACCAGAGAAATGAAGCGCA 1097
Db 1241 AGCCGGCCGAGAGAGATTTACTTGGAAACCTGGAAAGCTTGTCTACACAAAGCGGGCG 1300
Qy 1098 GCCCGCTGGCGAGAGAGAGCGGCTCGCTGTGCTGCGCTCTCGCCGGAATGTCGGA 1157
Db 1301 AATGCACTGCGCCCGGAGACACAGTCAACCGGTCTGCTTCCAAACGCAAGAGGTTCCGAG 1360
Qy 1158 AGCATCTGCGCCGGGTATAGGCTGTAGCGGCAAGCCC---ATCTACATCAGCGAAGC 1214
Db 1361 AGCTGCTGTTTGGCTCAGCAAGCGTTTCACTAACCCCAAGATCTTGTTACGAGAAATG 1420
Qy 1215 GATGCCCGTCCCTGAGAGAGAAACATGACGTGCGAGAGAGCGCGTCAACGACCCCTTC 1274
Db 1421 GCAAGTGTGTCMAAGGGGAGAACGATATGCCGCTCGAGAAAGATCTCGAGATGACTTCC 1480
Qy 1275 GCATCGSTACTTTGACTCGCACTGGACTGATTTTCAAGGCACTTATCCAGAGCGCG 1334
Db 1481 GGGTCCAGTCTACAGATTTATGTCAAGGCTTTGGCGCGGCTTATTCGAGGATGTG 1540
Qy 1335 TCGTGTCAAGGGGTACTTTGCGGGCGTTGCTGATCTGATTAATGGAATGTCAGATGCT 1394
Db 1541 TAAACGTCCCGGGTACTCAGCTTGTGCTGTGATGACAACTTGAATGGGCGAGGGCT 1600
Qy 1395 ACAGACCAAGATTGCGGCTCAGCTTCAACAGACTAC--ACCACCTTCAAGCGACGCCCA 1451
Db 1601 ACAGAGACCAAGTTTGGTGTGCTACCTTTGTGACTACAGAAACGCGCAGAACTTACCCA 1660
Qy 1452 AGAAGTGTGCTGTGCTTCAAGACATGTTTG 1484
Db 1661 AGAAGAGCGCCAAAGCAATGAAGCCGTTGTTTG 1693

```

## RESULT 5

AY109400 2031 bp mRNA 1linear HTC 17-OCT-2002

LOCUS AY109400 Zea mays CL1052\_1 mRNA sequence.

DEFINITION Zea mays CL1052\_1 mRNA sequence.

ACCESSION AY109400 GI:21213111

VERSION AY109400.1 GI:21213111

KEYWORDS HTC.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

1 (bases 1 to 2031)

Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitelc, M.S.,

Arthur, L.W., Henafey, M., MORGAN, M., and TINGEY, S.V.

Maize Mapping Project/Dupont Consensus Sequences for Design of

Overgo Probes

Unpublished (2002)

2 (bases 1 to 2031)

Coe, E.H.

Direct Submission

Submitted (25-APR-2002) Maize Mapping Project, University of

Missouri, Columbia, MO 65211, USA

These are publicly available in getting corresponding physical clones,

searching at MSL, maize-map.org; ZMB, www.zmbd.iastate.edu; TIGR,

www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the

maize cDNA sequences is either Virginia Walbot, Stanford or Pat

Schnable, Iowa State, then clones may be requested from ZMB:

www.zmbd.iastate.edu.

location/Qualifiers

1..2031

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="MaizeDB:629921"  
 /db\_xref="taxon:4577"  
 /clone\_id="Maize Mapping Project/Dupont Cornsensus  
 Library"  
 /note="this sequence is part of a project of EST  
 assemblies resulting from the application of public  
 configs to seed Dupont contigs; this resource was  
 assembled by Dupont as part of a collaboration for the  
 overgo addressing of BACs in conjunction with the Maize  
 Mapping Project"

## ORIGIN

Query Match 10.1%; Score 201; DB 11; Length 2031;  
 Best Local Similarity 51.3%; Pred. No. 1.3e-25;  
 Matches 705; Conservative 1; Mismatches 632; Indels 36; Gaps 9;

116 GCCTACAGATCGAAGGCGCGCTCAAGAAAGAGTGGCGCGCGCGCGCTCCATCTGGAGACG 175  
 5 GCGTACCGAGTTACGAGGCGCGCTCCACCAACGCGCGCGCGCGCTCCATCTGGAGATTCA 64  
 176 TACGCGACCTGAGCCATCGCGACCAACGCGCGCGCGCTGGGCTTGGGATCAC 235  
 65 TTGCGGACGCTCCAGGAAATATTGACAGGAATCAAAATGGAGAGCTTGAGTGATCA 124  
 236 TACCACCGCTACGATGAGAGCTTGTATCTTTGACCAAGTACGGCGAAAGGCGCTTACCGC 295  
 125 TACCATGCTACAGAAAGAGTGCATCTCATGAAAAGTTGAACTTTGATGCTTACCGG 184  
 296 TTCTCTTGTGCTGCGGATCATTCCTCCGCGCGAGGCTGATCCGTCACGAG 355  
 185 TTCTCATCTCATGTCGAGGATCTCCGAGTGGCGAAGG-----AAAGTCAATCCA 238  
 356 GAGGGAATTAGTTTACAGAACTGATGACGCGCTTGGAGCGCGGATACACCGCT 415  
 239 GAGGTTAGGTTATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 298  
 416 TGGGTGACTTTGACCACTGGATCTGCTCAGCGGCTTCCAGTCCGATGAGAGCTGG 475  
 299 TACATCAACCTTTACCACTATGATCTTCTTGGCTTGAGAAATATGAGAGGTTGG 358  
 476 CTCAACGTGAAAGGTCAGCTGAGCTTGGAGCGGTATGCGAGTTGTCTTTGAAAGT 535  
 359 TTAAAGCCGGAAGTGGCGGACTTG---TTTACAGACTATGCTGACTTCTGTTTAAAGC 415  
 536 TTTGGGAGACCGAGTCCAGAACTGATCAACATCAACGACCTTGATTCAGGCACTAT 595  
 416 TACGCGCATGCGTAAAGCACTGGTTTACATTAATGAGCCAAAGATATGAGCGCTACTT 475  
 596 GGAATATGCCACCGGACCAACGCGCGGACGAGACGATTAACAAGCACTCCACCGAG 655  
 476 GGCATATACACAGGTCAAATCTCTCTCAAAAGTGCACAGATGTG-----CTGCTGGT 529  
 656 GCGAAGCTGCACTGAGCGCTGCTCGCTGAAAGGCGCGAGATCATGAGCCATGCCCGC 715  
 530 GGGAAATTCAGGAACGGAACCTTATATGTTGCTCATATATTCTTGCGGACATGTAAT 589  
 716 GCGGTGCGGCTCTACAGAGGATTTTGCGCCCTCGCAAAAGGCGCGAGATGGCATCTCG 775  
 590 GCAATTCAGAAATACCGTACAAATATCAGGCTCTCAGAAAGGTTAAGTGGTGAATGTC 649  
 776 CTCACGCGGCTCTACTATGAGCCCTGGGACAGCATGAGCTCGGACCAAGAGGCTGCT 835  
 650 CTGGACTTTCAACTGTTACGAGG---TCTTACAAATCACCCTGATGACCAAGACGAGCC 706  
 836 GAGCGACGATGAAATTTACATTTGCTGTTTCCCAATCCCATCTTTTGAAGAAGAC 895  
 707 CAAGAGACCGAGGACTTCCATTTGCTGTTTNTATTC---AGCAGAGACGAGAC 763  
 896 TATCAGAGAGCAAGAAGAGCACTGGCGAGAGGCTTCAGCGCTCACTCCGCGGAC 955  
 764 TATCAGAGATTAATGACAGATCTCTGAAGAGGAGCGCCAGGTTTCACTTGAGCAG 823  
 956 TTTGCCATCTCAATGCGGAGAGACGACTTCTACGCGATGAAATTAC---TACACATCC 1012

Db 824 GCTAACTGATGAAGGCTCGGACGACTACATCGATCAACGAGTACACATCCAGCTAC 883  
 Oy 1013 CAGTTCGCGGCCACCTAAGAGGATCCGCTCCCGAAGAGGACATATCTCGGCGCATCAT 1072  
 Db 884 ATGAGGAGGAGAGAGTGTGTCAGCTGGCGCCAGTAGTACTTCCGATTGGCAGTT 943  
 Oy 1073 GAGCAGCAGAGAGATTAAGACGCGAGCGGCTTGGCGAGAGAGAGCGCTCGCTGCTG 1132  
 Db 944 CAATATGACATGCGACCGCAATGGCAACCATTTGACCAAGCGAATTTCTAGTGGCTC 1003  
 Oy 1133 CGCTCTGCGCGGACATGTTCCGAAAGCATCTGCGCGGCTGTAAGGCTGTACGGAAG 1192  
 Db 1004 TACATCGCCCCGACGAGGATGATGAGGTGGTGAATCACTCAAGAGAGATGAGGAAT 1063  
 Oy 1193 C---CCATCTACATCAACCGAAGACGATGCGCGCTCGCTGAGAGAGAACATGACGTC 1249  
 Db 1064 CCAACGATCTACATTAAGGAGAAACGAAATGACCAAGCAGCTTGA-----AACTGACCGA 1117  
 Oy 1250 GAGGAGCGCTCAACGACCGCTTCCGATCCGATCGGATCTTGAATCGCACTTGAATGAT 1309  
 Db 1118 GACCATGACTCGGCGGACGCGACGAGAGGTCGCTTACAGAGCTACATCGCGCAGCTG 1177  
 Oy 1310 TCCAGGCTATTAACCCAGACGAGCGCTGCTGCAAGGAGTACTTGGCGGCGTTCCTC 1369  
 Db 1178 AGAAGGCTATGACGAGGAGCGAAGCGAAGCTGGC---TGGCTACTTGGCTGCTCTCTC 1234  
 Oy 1370 GATPACTGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1429  
 Db 1235 GACAACTTCAGATGCTGCTGAGGAGTACTGCTCAAGTTGCGCATCTGCTAGGAGCTTC 1294  
 Oy 1430 ACCACCTCAAGGCGACGCCAAGAAAGTCTGCGCTGCTCTCAAGGACATGTT 1483  
 Db 1295 AACAGCTTCAAGACCCACCCGAAAGGCTCGGCTTACTGTTCAAGGACATGCTT 1348

RESULT 6  
 CB677158 788 bp mRNA linear EST 09-APR-2003  
 LOCUS OSUNBEI3N03.f OSUNBE Oryza sativa (japonica cultivar-group) cDNA  
 DEFINITION  
 Clone OSUNBEI3N03 5', mRNA sequence.  
 ACCESSION  
 CB677158  
 VERSION  
 CB677158.1 GI:29680883  
 KEYWORDS  
 EST.  
 SOURCE  
 Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE  
 1 (bases 1 to 788)  
 Jantsuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 JOURNAL  
 COMMENT  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aac cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 13 row: N column: 03  
 Seq primer: gta aac cga cgg cca gtc.  
 Location/Qualifiers

FEATURES  
 source  
 1..788  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"

```

/db_xref="taxon:39947"
/clone="OSJNBE13N03"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNBE"
/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after immunoliation with Rice Blast (70-15)"

```

## ORIGIN

```

Query Match      9.8%; Score 195.2; DB 14; Length 786;
Best Local Similarity 56.3%; Pred. No. 1.8e-24;
Matches 409; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

QY 81 TCCCGAACGACTTTAATGAGGCTTCGCAACGCGCGCTACAGATTCGAGGCGCGCTCA 140
DB 60 TCCCGAGAGACTTCACTTCGCGACCGCTCGCGCTTATATAGACGAGGCGCTTGA 119
QY 141 AAGAAAGTGGCCGCGGCGCTTCATCTGGGACACGTACTGCCACCTGAGCCATGCGCA 200
DB 120 ACGAAGCGCGGACGAGCGCGGACATCTGGGACACGTAACGTAATCCAGGCAAGTTG 179
QY 201 CCAAGCGGCGGACGAGCGGATGCTTGATCAACCGCTACAGATTCGATGAGACTTTG 260
DB 180 AAGATGTAAGCAAGCGGAGTGAAGATGATTTTACATGCTACAGAGAGATCTGA 239
QY 261 ATCTCTTGAACCAAGTACGCGCGGCAAGGCTTACCGCTTCTCTGTCGTCGCGATCA 320
DB 240 ACTTCGTACCGACATGAAATGACGCGCTTCCTTCATTCCTTTGGAGCAGAGATCC 299
QY 321 TTTCCCTCGGCGGCGAGCTGATCCCGTCAACGAGAGGGAATTGATTTACAGCAAAAC 380
DB 300 TCCCAAAATGAAACCATCAGTGAAGGAATCAACAAAGAGGATGCTTTCAACAAACCC 359
QY 381 TGATGACGCGCTGTTGAGCGGCGGTATCAAGCTTGGGTGATCTTGATCCATGCGGATC 440
DB 360 TGATTAAGAGGTATCTATAGAGGCTTGAAGCCATTTGTCACTTTCATTTGCA 419
QY 441 TGCTCAGCGCTTACGATGCTATGAGAGGCTGCTCAACGTGGAAGAGGTCCAGCTG 500
DB 420 CCCCACAGGCTCTGAGAGCAAAATACCGAGCTTCTCA---GTGAAAACATTTGAAG 476
QY 501 ACTTGACGCGGTATCCGAGGTGTTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTGA 560
DB 477 ATTTCGTGACTATGCGAGCGTGTCTTCGTGATTCGGGACCGGCTGAAGTATGGA 536
QY 561 TCACCATCAACGAMCCCTGATTCAGGCTATGATATGATATGCCACCGGACCAACGCC 620
DB 537 ACAAGTTCAAGAGCCGATGATCTTCTGCGCGCGGCTTACGCTCCGACCAAGGCC 596
QY 621 CGGCGAGAGCAG-----CATTAACAAGCACTCCACCGAGGGAACACTGCCACTGAAC 674
DB 597 CCGGCGCGCTGCTCGCGCTGACGCTTCACAAAGATCGCCCCCGGCGACTCCGCAACG 656
QY 675 CGTGCGCTGCTGAAAGGCCAGATCATGAGCCATGAGCCCGCGCTGCGCTTACAGCA 734
DB 657 CGTACGTGCGCGGCAACAACCTGCTGCTCGCCACGCGGAGCGCTCGCTTACCGCC 716
QY 735 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCTATCGCTCAACGCGCAGTACTATG 794
DB 717 AAAAGTACGAGCGGACGAGAAAGGGGAGATCGGCTATGAGCAGAGTGTGCACTGTTG 776
QY 795 AGCCCT 800
DB 777 TGCCCT 782

```

```

RESULT 7
LOCUS CB673495 786 bp mRNA linear EST 09-Apr-2003
DEFINITION OSJNBE08A17.1 f OSJNBE Oryza sativa (japonica cultivar-group) cDNA
ACCESSION CB673495

```

```

VERSION CB673495.1 GI:29677220
KEYWORDS EST
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretidae; Oryzae; Oryza.
REFERENCE Jantaauriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Ming,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
AUTHORS Contact: Rod Ming
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

```

## JOURNAL

## FEATURES

```

PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: A column: 17
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1..786

```

## ORIGIN

```

Query Match      9.7%; Score 193; DB 14; Length 786;
Best Local Similarity 57.1%; Pred. No. 4.4e-24;
Matches 437; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

QY 127 CGAAGCGCGCTGAAAGAGGTGCGCGCGCTTCATCTGGGACAGTACCGCACT 186
DB 22 CAGGAGCTGTGAAGAGAGCGGAGAGGCGACATCTGGGACAGCTTGGCGCAC 81
QY 187 GGAAGCATCGCGACCAACGCGCGCAACGCGATGTGCTTGGATCACTACCAACCTA 246
DB 82 CTTGGAAGATCACGCACTTCAAGCATGTGATGTGAGTTGATCACTACCAACCTT 141
QY 247 CGATGAGACTTGATCTTGAACCAAGTACCGCGCAAGGCGCTTCCCTTCTTGTG 306
DB 142 CGAGGAGATATACACTATGACAGATGAGGATGATGTATCGGCTTCTGATAC 201
QY 307 GTGTGCGGATCATTTCCCTCGCGCGGAGCTGATCCGCTCAACGAGAGGAATTGA 366
DB 202 ATGTGTAAGATTTACCAATGTG-----TTGATCAAGTCAATCAAGCTGTGATCA 255
QY 367 GTTTTACAGCAACATGATTGACGCGCTGTGAGCGCGGATTCACGCTTGGGTGACTT 426
DB 256 CCACTACAAACAAGCTATGATGATGATCTTACCAAAAGAAATTGACGCAATGTGCACT 315
QY 427 GTACCATGAGATCTGCTTCAAGCGCTTACAGATGCTATGAGAGTGTGCTCAAGTGA 486
DB 316 CTACCATGAGACTTCCCGAGCGCTTGAAGACAAAGTCAAGGCGTGTGAGAG-- 373
QY 487 AGAGGTCCAGTGAATTTGAGCGGATGCGAGTGTGCTTGAACGTTTGGGAGCG 546
DB 374 -CAGATAGTGAAGATTTGCGCGCGTACCGGAGACGTCTTCAAGGAGTTCGGGAGCAG 432

```

QY 547 AGTCCAGACTGATCCATCAACGAMCCCTGATTACAGCCATCTATGATATGCCAC 606  
 Db 433 GGTGAACACTGATGATCAAGCTCAACGAGCGCACACGCTGATCCAGGGCTACGACGC 492  
 QY 607 CGGCAGCAAGCCCCGGGGAGAGCAGCATTA-----ACAAGCACTCCACCGAGGCCAA 660  
 Db 493 AGGAGCTCAAGCCCCCGGCGCTGCTCCGCTGCTCCACCTCTACGCAAGGCCGCGCAA 552  
 QY 661 CACTGCACATGAGCCCGTGGCTGCTGGAAGGCCACAGATCATGAGCCATGCGCGCGCT 720  
 Db 553 CTCGGGACCGAGACCTTACGTCGTCTGCGCCACCACTTATCTGCGCCACGCGCGCGC 612  
 QY 721 GCGCGTCTACAGAGGAGACTTTGCGCCCTCGCAAAAGGCGCAGATCGCATCTGCTCAA 780  
 Db 613 CACACTCTACAGAACAAATACAAAGGCGACGAGAACGGGACGCTTGGATAGGTTTGA 672  
 QY 781 CGCGCACTACTATGAGCGCTTGGAGACGATGAGCCTCGGAGCAAGAGCTGTGAGCG 840  
 Db 673 CGTGAATGTGGTGGAGCCGATGTCACAA---CCACATGACATCGAGCGCGCAAGAG 729  
 QY 841 ACGGATGAATTTACATGCTGCTGTTTGCATATCCCATCTTCTT 885  
 Db 730 AGCGCAGGAGTTTCACTAGAGATGTTTGTCTGATCCGTTCTT 774

RESULT 8 CB642881 806 bp mRNA linear EST 08-Apr-2003  
 LOCUS OSJNB03E17.f OSJNB Oryza sativa (japonica cultivar-group) cDNA  
 DEFINITION Oryza sativa (japonica cultivar-group) EST.  
 ACCESSION CB642881 GI:29637872  
 VERSION CB642881.1  
 KEYWORDS Oryza sativa (japonica cultivar-group)  
 SOURCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 ORGANISM Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 EST. 1 (bases 1 to 806)  
 Jantschuliyarat, C., Lu, G., Gowda, M., Harfield, J., Zhou, B., Mazur, E.,  
 Kudirna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 JOURNAL Contact: Rod Wing  
 COMMENT Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR PRIMERs  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tac gac cat g  
 Plate: 03 Row: B Column: 17  
 Seg primer: gta aaa cga cgg cca gtc.  
 FEATURES  
 source  
 1..806  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /molecule="mRNA"  
 /db\_xref="taxon:39947"  
 /clone="OSJNB03E17"  
 /issue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_11b="OSJNB"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2:  
 XhoI, 24 hrs after inoculation with Rice Blast (che  
 86061)"

ORIGIN

Query Match 9.6%; Score 190.4; DB 14; Length 806;  
 Best Local Similarity 57.5%; Pred. No. 1,3e-23;  
 Matches 407; Conservative 0; Mismatches 286; Indels 15; Gaps 3;

QY 81 TGGCCAAACACTTTGATGGGCTTGGCAAGCGCGCTTACCGAGTGAAGGCGCGCTCA 140  
 Db 96 TCCCGAGGGGTTGCTTTCGCGACCGCTCCCGCGTACCGATGAGAGGAGCTGTGA 155  
 QY 141 AAGAGGTGGCGCGCGCGCTCCATCTGGGACAGTACTCCACCTGAGACCAATCGCGCA 200  
 Db 156 AGAGAGACGGAGAGGAGGACAGCATCTGGGACAGCTTGGGACACTTTGGAAAGATCA 215  
 QY 201 CCAAGCGCGCAACGAGATGTGGCTTGGCATCTACCAACGCTAGATGAGACTTTG 260  
 Db 216 CGAGCTTACAGCAATGCTGATGTTGATGATGATGATGATGATGATGATGATGATGAT 275  
 QY 261 ATCTCTTACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTGTGTGTGTGTGTGTG 320  
 Db 276 AGCTCATGGCAGACATGGGGATGGATGGCTATGCTTCTCGATAGCATGTCTAAAGATCT 335  
 QY 321 TTCCCTCGCGCGCAGGCTGATCCGCTCAAGAGGAGGGAATTTGATTACAGCAAC 380  
 Db 336 ACCCAATAGGTG-----TTGTCAAGTCAATCAAGCTGTATGACACTTACCAAGAC 389  
 QY 381 TGATGACGCGCTTGTGAGCGGGGTATACAGCTTGTGATCTTGTATCACTGAGATC 440  
 Db 390 TGATGATGACTTCTAGCAAAAGAAATCAGCATATGATGATGATGATGATGATGATGATG 449  
 QY 441 TGCTTACGCGCTTTCACGATCGTATGAGAGCTGTGCTCAAGTGAAGAGTCCAGCTGG 500  
 Db 450 TCCCGCAGCGCTTGAAGACAGTACAGAGGCTGTGTCAGAG---CAATAGTGGACG 506  
 QY 501 ACTTGAAGGGGTATGAGAGGTTGTGCTTGAACGTTTGGGGACCGAGTCCAGAACTGGA 560  
 Db 507 ATTTCGGCGGTATCGGAGACGCTTTCAGGAGTTCGGGAGCAGGTAAAGCATCTGA 566  
 QY 561 TCACCATCAACGAMCCCTGATTCAGGCCATCTATGATATGTCACCGCAGCAACGCC 620  
 Db 567 TCAAGCTCAACGAGCGCGCACACAGTGGCCATCCAGGGCTACGAGCAGGCTCCAGGCC 626  
 QY 621 CGGCGAGAGAGCAGC-----ATTACAGACCTCCACCGAGGCAACACTGCTGAGC 674  
 Db 627 CCGCGCGCTGCTTCCGCTGCTCCACTTACCTGCAAGGCCGCAACTCCGCGCACCGAGC 686  
 QY 675 CGTGGCTCGCTGGAAGGCCACATGATGAGCCATCCCGCGCGCTGCTTACAGCA 734  
 Db 687 CTTAGCTGTCGCGCCACCACTTCTGCTCCACCGCGCGCGCGCGCATTTACAGGA 746  
 QY 735 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGCATCTGCTCAAG 782  
 Db 747 CAABAATCAAGCGAGCGCAAGACGGGCACTTGGATAGGTTTCAAG 794

RESULT 9 CD438244 889 bp mRNA linear EST 03-JUN-2003  
 LOCUS EL01N0511A05.b Endosperm\_5 Zea mays cDNA, mRNA sequence.  
 DEFINITION CD438244  
 ACCESSION CD438244.1 GI:31353887  
 VERSION CD438244.1  
 KEYWORDS EST.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogonae; Zea.  
 1 (bases 1 to 889)  
 Lai, J., Dey, N., Kim, C.S., Becraft, P., Larikins, B., Linton, E. and  
 Messing, J.  
 Sequencing of the maize endosperm ESTs  
 Unpublished (2002)  
 JOURNAL Contact: Lai, Jinseng  
 COMMENT Dr. Joachim Messing's lab  
 Wakeman Institute, Rutgers University

190 Prellinghuysen Rd., Piscataway, NJ 08854, USA

Tel: 732-445-3801

Fax: 732-445-5735

Email: j1a1@akman.rutgers.edu

Seq primer: T3.

# FEATURES

source

Location/Qualifiers

1. 889

/organism="Zea mays"

/mol\_type="mRNA"

/cultivar="W22"

/db\_xref="taxon:4577"

/issue\_type="Endospore of 7-23DAP"

/clone\_lib="Endospore 5"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 9.0%; Score 179.2; DB 14; Length 889;  
Best Local Similarity 54.8%; Pred. No. 1.2e-21;  
Matches 493; Conservative 0; Mismatches 383; Indels 24; Gaps 6;

```

QY 113 GCCGCTACAGATCGAAGGCGCGCTCAAGAGGTGCGCGCGCTTCATCTGGGAC 172
Db 5 GCGGTATCCAGATCGAAGGCGCGCGCTCAAGAGGTGCGCGCGCTTCATCTGGGAT 64
QY 173 ACCTACTGCACTGAGGACATCGGCAACCAAGCGGCAACGGGATGCTTGGGAT 232
Db 65 TCATTCGCGCAGTCCAGAAATATTCAGGAAATCAAAATGAGAGCTTGGACTGAT 124
QY 223 CACTACCAAGCTACGATGAGGACTTGTATCTCTTGCACCAAGTACGCGCAAGGCTTAC 292
Db 125 CAATTCATTCGCTCAAGGAAAGACCTGCATCTCAAGAAAGTTGAATCTTGATGCTTAC 184
QY 293 CGCTTCCTGCTGCTGCTGCGGATCATTCCTCGCGGCGGAGCTGATCCGCTCAC 352
Db 185 CGGTCTCAATCTCATGTGCTCCAGATCTTCCGAGTGGCGAAGG-----AAATCAT 238
QY 353 GAGAGGGAATGATTTTACGCAAACTGATGCGGCTTGGAGCGGGGATATCAG 412
Db 229 CCAGAGGTGATGCTTATTAATATTTGATTAATCTTCTGCTTCAAGAGGCTGATCT 298
QY 413 CTTGGGTGACTTTGTACCACTGAGATGCTGCTCAGGCGCTTACGATCGCTATGAGGC 472
Db 299 CCTTACATCAACCTTACCACTATGATCTTCTCTTGGCTTGAAGAATATGAGAGG 358
QY 473 TGGCTCAACGTGGAAGAGTCCAGCTGACTTTGAGCGGTATGCGAGGTGTGCTTGA 532
Db 359 TGGTAAAGCGGAGATGCGGAGCTTG---TTTACAGACTATGACTTCTGTTTAA 415
QY 533 CGTTTGGGAGCGAGTCCAGAACTGATCACTACCAAGCAAGCCTGATTCAGGCCATC 592
Db 416 ACCTACGCGGATCGGTAAGACATGCTTTTACATTCATGAGCCAGAGATGATGAGCT 475
QY 593 TATGATATGCAACCGGAGCAAGCGCCCGGAGAGAGAGCATTAACAGACTTCCACC 652
Db 476 CTGGCTATGACACAGGATCAATCTCTCTCAAGAGTG-----ACAGATGCGCTGCT 529
QY 653 GAGGCAACACTGCACTGAGCGGCTGCTGCTGAGAAAGGCCAATATGAGCCATGCC 712
Db 530 GGTGGGAATTCAGCAACCGAACCTTAACATATGTTGCTCAATATTTTCTTGGCAATGT 589
QY 713 GCGGCGTGGCGCTGACAGAGGAGCTTGGCCCTCGCAAAAGGCGCAGATCGGCAATC 772
Db 590 ACTGCACTTTCAGAGATACCTGACGAATATCAGGCTGCTCAGAGGATTAAGTCCGATA 649
QY 773 TCGCTCAACGCGACTACTATGAGCGCTTGGAGCAGCAATGAGCTTGGAGCAAGAGCT 832
Db 650 GTCCTGAGATTCATCTGATGAGGCTTT---ACAACTCACCTGATGACCAAGAGCA 706
QY 833 GCTGAGCAGCGATGGAATTTTCACTTGGCTGTTTGGCAATCCATCTTCTTGAAGAG 892
Db 707 GCCCAAGAGCGAGGACTTCCATTTGCTGCTGTTTGTGATCC---ATTGATTAACGGA 763

```

```

QY 893 GACTATCCAGAGACATGAAGACAGCTGCGGAGAGGCTTCCAGCCCTCACTCCGCG 952
Db 764 CACTATCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 823
QY 953 GACTTGGCACTCTCAATGCGGAGAGACCGACTTCTACGCGCATGATTACTACATCC 1012
Db 824 CA---GGCTAACTGATGAGGCTGCGGAGACTACATCGTATCAACGATACATCC 880

```

## RESULT 10

CB688082

LOCUS

DEFINITION

CB688082 480 bp mRNA linear EST 04-SEP-2003

CB688082.1 GI:34448951

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

CONTACT: Dave M., Nuss DL

UNIVERSITY OF MARYLAND BIOTECHNOLOGY INSTITUTE

5115 PLANT SCIENCES BUILDING, COLLEGE PARK, MD 20742, USA

TEL: 301-405-7661 or 5111

FAX: 301-314-9075

EMAIL: dawem@umd.edu

Cryptonecristia parasitica EST derived from mixed samples of cDNA

from wild-type (strain EP155) and hypovirus-infected cultures.

Seq primer: M13-reverse.

Location/Qualifiers

1. 480

/organism="Cryptonecristia parasitica"

/mol\_type="mRNA"

/strain="EP155"

/db\_xref="ATCC (inhost):38755 and 52571"

/db\_xref="taxon:5116"

/clone="EP155, EP155-CHV1-EP713"

/clone\_lib="Mixed source, strain EP155 and EP155 infected

with hypovirus CHV1-EP713"

/note="Cryptonecristia parasitica EST derived from mixed

samples of cDNA from wild-type (strain EP155) and

hypovirus-infected cultures"

## ORIGIN

Query Match 8.8%; Score 175.6; DB 14; Length 480;  
Best Local Similarity 64.4%; Pred. No. 6e-21;  
Matches 290; Conservative 0; Mismatches 158; Indels 2; Gaps 2;

```

QY 72 CGTAGCTTCCCAAGCACTTTGATGAGGCTTTCGCAAGCGCGCTTACCAAGTGAAG 131
Db 4 CGGTCGCTACATCAAGACTTCAAGTGGGCTTTGGCCAGCGGCTCTTCAACAGGTCGAAG 63
QY 132 GCGCGCTAAAGAGAGTGGCGCGCGCGCTTCACTTGGGAGACAGATCGCAACCTGAGC 191
Db 64 GCTTACCGAAGAGAGCGCGCGCGCGCTTCACTTGGGAGACAGATCGCAACCTGAGC 123
QY 192 CATGCGCAACCAAGCGCGCGCGGATGTGGCTTGGGATGATCACTACCAAGCTTACGATG 251
Db 124 GCAAGATGCGGAGCGGCTTCAAGCGGCGGCTGCTGAGACAGTACCAAGCGCAAGCGCG 183

```



OY	252	AGGATTTTGATCTCTTGACCAAGTACGGCGCAAAAGCCTTACCGCTTCTCTGTGTGAGT	311
Db	184	AGGATTTGCGCTGCTCAAGACGCTCGGGGCGACGGCTTACCGTTCTTCGCTGTGTGTGT	243
OY	312	CGCGGATATTTCCCTCTCGGCGGACAGGCTGATATCCCGTCAACGAGAGGGAAATTGATTTT	371
Db	244	CGCGGATATATCCCGTGTGGCGGGCGCGACGACCCCGTCAACCAAAAGGGGCTGGACCACT	303
OY	372	ACAGCAAACTGATTTGACGCCCTGTTGAGCGGGGATATCAGCGCTTGCGGTGACTTTGTATCC	431
Db	304	ACAGCGGCTTACCAACGACCTCTGCTGGCGCGGGATATAGCCCTTCAATCAACGCTGTACC	363
OY	432	-ACTGGGATCTGCCTCAGGCGGCTTACGATGGCTATGAGG-CTTGCTCAACGTGGAAGA	489
Db	364	AACGTGGGACTCGCCGACGAGNCTGGACCGGGGTACCGCGGCGCTGCAACCGCACNNA	423
OY	490	GGTCCAGCTGGACTTTGAGCGGTATGCGAG	519
Db	424	GTTCCCGCTTGACTTTGCCCGGTACGCCGC	453

[illegible]

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 1182)	Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes

JOURNAL Unpublished (2002)  
 REFERENCE 2 (bases 1 to 1182)  
 AUTHORS Coe, E. H.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of  
 Missouri, Columbia, MO 65211, USA  
 COMMENT If you are interested in getting corresponding physical clones,

**COMMENT**  
If you are interested in getting corresponding physical clones, these are publicly available from zmbd and may be found by BLAST searching at MSU, [maizemaps.org](http://maizemaps.org): zmbd, [www.zmbd.iastate.edu](http://www.zmbd.iastate.edu): TIGR, [www.tigr.org](http://www.tigr.org) or NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov). When the source of the maize cDNA sequences is either Virginia Waldbop, Stanford or Pat Schnable, Iowa State, then clones may be requested from zmbd: [www.zmbd.iastate.edu](http://www.zmbd.iastate.edu).

FEATURES	location/qualifiers
source	1. .1182

```

/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:635002"
/db_xref="taxon:4577"
/clone_id="Maize Mapping Project/Dupont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
configs to seed Dupont configs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

```

	ORIGIN
Query Match	8.8%; Score 175; DB 11; Length 1182;
Best Local Similarity	54.8%; Pred. No. 5,9e-21;
Matches 440; Conservative	0; Mismatches 343; Indels 18; Gaps 4;
DY	77 GCTCTGCCCAACGACTTGAATTGGGGCTTGCCAAAGCGCGCCTTACAGATCGAAGCGCC 136

Db	205	GC GTTCCCGGACGGCCTTCGTCCTTGCGGAGCGGCGCGTCGCGATCCACGAGTCGAGGGGATG	264
Oy	137	GTCAAAGAAGTGGCCCGCGCGCCCTCATCTTGGGACACGATCTGCCACCTGGAGCCATCG	196
Db	265	GCCAGAGACGGCGGGCGGGGCCCGCAGATCTGGGAGCGCTTCATAGAGTTCCCGGACC	324
Oy	197	CGCACCAACGGCGGCAACGGCGGATGTGGCTTTCGATCACTACCAACCCGCTACGATGAGAC	256
Db	325	ATCCCTTAACAATGCCACCGCTGACGTACCGGTGACGAGATATCATCGGTACAAAGGAAGAT	384
Oy	257	TTTGATCTCTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGCTGGTGC CGG	316
Db	385	GTGAACATATGAAAGAACATGGGCTTTGATGCGTACGGGTTTTCGATCTCTTGCTCCAGG	444
Oy	317	ATCATTCCTCCCGCGGCGGACGGCTGGATCCCGTCAACGAGAGGAGATTTGATTTTACGC	376
Db	445	ATTTTCCAGATGGAAGTGGCAAC-----GTGAACAAGAAAGGATGGATTTACTACAA	498
Oy	377	AAACTGATTTGACGCGCTGTTGAGCGCGGGGATATACGCGCTTGGGATCTTTGTAACACTGG	436
Db	499	AGGCTCATAGATTATCATGCTCCAGCAAGATATCGCGCGGTATGGAATCTCTACATTAT	558
Oy	437	GATTCGCTCAGCGGCTTACGATCGTATGAGAGCTGGCTCAACGTGGAAGAGCTCAG	496
Db	559	GACCTCCCATTTGGCACTCCATGAAACATAGTACTCGGGCTGGCTTAGCC--CAGAATTTGTG	615
Oy	497	CTGGAATTTGAGCGGATATGCGAGGTTGTGCTTTGAACGTTTTGGGGACCGAGTCCAGAC	556
Db	616	GAGCGGTTTGAGACTACGCGGAGTTCTGCTTCCACCGCTTCGGAGACAGGGTAAAGAAC	675
Oy	557	TGATATCAACATCAACGAMCCCTGATTCAGGCACTTATGATATGCCACCGCAGCAAC	616
Db	676	TGGTTTACTTTCAACGAGCGGAGGTGGTGGTCTCTGCGGCTTACGACATATGGCTTGAC	735
Oy	617	GCCTCGGAGAGAGACGATTTAACAGCACTCCACGAGGGCAACCTGCCACTGAGCCG	676
Db	736	GCACCGGGAAGGT-----TTCCGGGTGCGCCCGCGGAGGCAATCCACACACGGAACCG	789
Oy	677	TGGCTCCCTGGAAAGGCCAGATATAGGCAATGCGCGCGCGCGCGTCTAAGCAGG	736
Db	790	TACCTTGTGACACCAATCTCATCTTTTCTATGACGCTGGGTCAGGCGATACCGGAC	849
Oy	737	GACTTTGGCCCTCGCAAAAGGCGCAAGTCGCACTTCGCTCAACGGCGCATACTATGAG	796
Db	850	AAAGTATCAGCTTACACGAAGGGGAATTTGAAATTTCTCTGGAATTTCTGTGGTTAGAA	909
Oy	797	CCCTGGGACAGCAATGAGCTCGGGAACAAGAGGCTGTAAGCGACGATGGAATTTTCA	856
Db	910	CTTT----TCAGCGACAGCAATGCGGACGAGGTGCAACAGCGAGCCAGGACTTCCAC	966
Oy	857	ATTGGCTGTTTGGCAATCCAT	879
Db	967	CTAGAGCTGTCTTTCAGCCCAT	989

RESULT 12					
LOCUS	CF880605				
DEFINITION	CF880605	662 bp	mRNA	linear	EST 31-OCT-2003
ACCESSION	trich082xh20_b1	Tr.reesl mycelial culture,			Version 6 October 2003
VERSION	Hypocrea jecorina cDNA clone	trich082xh20,			mRNA sequence.
KEYWORDS	CF880605.1	GI:38135287			
SOURCE	Est.				
ORGANISM	Hypocrea jecorina (anamorph: Trichoderma reesei)				
	Hypocrea jecorina				
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
	Hypocreomycetidae; Hypocreales; Hypocercaceae; Hypocrea.				
REFERENCE	1 (bases 1 to 662)				
AUTHORS	Dienes, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,				
	Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and				
	Dean, R.A.				
TITLE	Analysis of the protein processing and secretion pathways in a				
	Trichoderma reesei EST dataset				

## JOURNAL Unpublished (2003)

CONTACT: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu

## FEATURES

Seq primer: Lt-F1 primer.  
Location/Qualifiers

1. 662  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="OM6a"  
/db\_xref="taxon:51453"  
/clone="tric082xh20"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: PREP3J; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## ORIGIN

Query Match 8.6%; Score 171.8; DB 14; Length 662;  
Best Local Similarity 59.4%; Pred. No. 2.6e-20;  
Matches 366; Conservative 0; Mismatches 238; Indels 12; Gaps 4;

```

QY 646 CTCACCGAGGCGAACACTGCGACGAGCCGTCGCTCGTGGAAAGGCCGATCATGAG 705
Db 52 CGCCCCCGCGCGAGACACCTCGAGCCCGTGACGCGCCCAACATCTGTGCG 111
QY 706 CCATGCCCCGCGCGGTGCGCTCTACAGCAGGAGACTT--TCGCCCTTCGCAAAAGGGCCA 762
Db 112 CCACGCGCGCGCGGTGCGGTACCGGAGAGACTTCAAGCCCGCAGCGCGCA 171
QY 763 GATCGGATCTCGCTCAACCGCGACTACTATGAGCCCTGGGACAGCAATGAGCTTCGGGA 822
Db 172 GATCGGATCTCGCTCAACCGCGACTACTATGAGCCCTGGGACAGCGCGCGCA 231
QY 823 CAAGAGGCTGCTGAGCAGCGAGATTTCACTTGGCTGGTTCATCCCATCTT 882
Db 232 CAAGAGGCTGCTGAGCAGCGAGATTTCACTTGGCTGGTTCATCCCATCTT 291
QY 883 CTGAAAGAGACTATCCAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 942
Db 292 CTT---GGGCGACTACCGCGGTGATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 348
QY 943 CACTCCCGCGAGCTTGGCATCTTCATGTCGCGAGAGACCGACTTTTACGCGCATGAAATTA 1002
Db 349 TACGCGCGAGAGAGCGCGCTCTGTC---CGGCTCCAAAGCACTTTTACGCGCATGAAACA 405
QY 1003 CTACACATCCCATGTCGCGCGCACTAGACGATCCCGTCCCGAGACGACTATCTCCG 1062
Db 406 CTACACATCCCATGTCGCGCGCACTAGACGATCCCGTCCCGAGACGACTATCTCCG 465
QY 1063 CGCATTCATAGACACAGAGAAATAAGAGCGGAGCGCGCTTTGGCGAGAGAGAGAGAG 1122
Db 466 CAAGGTGACGTCGCTTTCACCAACAAAGCAGGCACTGATCGGCCCGAGAGAGAGAG 525
QY 1123 CGCGTGGTGGCTGCTGCGCGCGAGCATGTTCCGGAAGAGATGCGCGCGGTGTCAGGCTT 1182
Db 526 CCGCTGGTGGCTGCTGCGCGCGAGCATGTTCCGGAAGAGATGCGCGCGGTGTCAGGCTT 585
QY 1183 GTACGCGAA---GCCCATCTACATACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
Db 586 GTACGCGTACCGCGCGCATCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
QY 1240 CATGACGTGAGAGAG 1255
Db 646 CTTGCCAAGAGAGAG 661

```

## RESULT 13

CB907493

LOCUS

CB907493 752 bp mRNA linear EST 02-JUL-2003

DEFINITION

tric082xh20 T.reesei mycelial culture, Version 3 apr-11 Hypocrea

ACCESSION

CB907493

VERSION

CB907493.1 GI:30122151

KEYWORDS

EST.

SOURCE

Hypocrea jecorina (anamorph: Trichoderma reesei)

ORGANISM

Hypocrea jecorina

REFERENCE

Hypocrea jecorina

AUTHORS

Dunn-Colman, N.S., Goedgebuer, F., Houfek, T.D., England, G.J.,

TITLE

Transcriptional regulation of biomass-degrading enzymes in the

JOURNAL

J. Biol. Chem. 278 (34), 31988-31997 (2003)

MEDLINE

22603314

PUBMED

12788920

COMMENT

Contact: Pamela K. Foreman

Genencor Int'l.

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genencor.com

Seq primer: Lt-F1 primer.

Location/Qualifiers

1. 752

FEATURES

/organism="Hypocrea jecorina"

source

/mol\_type="mRNA"

/strain="OM6a"

/db\_xref="taxon:51453"

/clone="tric082xh20"

/dev\_stage="mycelia"

/note="Vector: PREP3J; Site 1: Not I/Sal I; Mycelial

culture grown from 24 hrs to 6 days with varying Carbon

and Nitrogen sources and concentrations."

ORIGIN

Query Match

8.6%; Score 171.8; DB 14; Length 752;

Best Local Similarity

59.4%; Pred. No. 2.5e-20;

Matches 366; Conservative

0; Mismatches 238; Indels 12; Gaps 4;

```

QY 646 CTCACCGAGGCGAACACTGCGACGAGCCGTCGCTCGTGGAAAGGCCGATCATGAG 705
Db 142 CGCCCCCGCGCGAGACACCTCGAGCCCGTGACGCGCCCAACATCTGTGCG 201
QY 706 CCATGCCCCGCGCGGTGCGCTCTACAGCAGGAGACTT--TCGCCCTTCGCAAAAGGGCCA 762
Db 202 CCACGCGCGCGCGGTGCGGTACCGGAGAGACTTCAAGCCCGCAGCGCGCA 261
QY 763 GATCGGATCTCGCTCAACCGCGACTACTATGAGCCCTGGGACAGCAATGAGCTTCGGGA 822
Db 262 GATCGGATCTCGCTCAACCGCGACTACTATGAGCCCTGGGACAGCGCGCGCA 321
QY 823 CAAGAGGCTGCTGAGCAGCGAGATTTCACTTGGCTGGTTCATCCCATCTT 882
Db 322 CAAGAGGCTGCTGAGCAGCGAGATTTCACTTGGCTGGTTCATCCCATCTT 381
QY 883 CTGAAAGAGACTATCCAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 942
Db 382 CTT---GGGCGACTACCGCGGTGATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
QY 943 CACTCCCGCGAGCTTGGCATCTTCATGTCGCGAGAGACCGACTTTTACGCGCATGAAATTA 1002
Db 439 TACGCGCGAGAGAGCGCGCTCTGTC---CGGCTCCAAAGCACTTTTACGCGCATGAAACA 495
QY 1003 CTACACATCCCATGTCGCGCGCACTAGACGATCCCGTCCCGAGAGAGAGACTATCTCG 1062

```

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Medline	PubMed	Comment	Features	Source
Db	496	CTACAGCTGCAACTACATCCGACACCGAGCTCCGCGCTCCGCGCGACGACACCGCTGG	555									
Qy	1063	CGCCATCCATGAGACCCAGAGATAGGACGGCAGCCCGTTGGCGAGAGAGCGCCTT	1122									
Db	556	CAACCTCGACGCTCTTACCAACAGACGAGGGCAACTGCACTGGCCCCCGAGCGCACTG	615									
Qy	1123	GGCTGGGTGGGCTCTCCGCGCGGACATGTTCCGGAAGCATCTCCGCGCGGGTGTACGGCT	1102									
Db	616	CCCCGTGGTGGCCCCCTCTGTGCGCGCGGCTTCCGCGCATCTTCTGTGTATCAGNCAGAG	675									
Qy	1183	GTACGCGCA---GGCCATCTACATCACCGAGACGATGCGCCGTGCGCGAGAGAGAA	1239									
Db	676	GTACGCGCAACCCCGCCCATCTACGTACGAGAGACGCAACGCACTCAAGGCGAGAGCA	735									
Qy	1240	CATGACGTGCGAGAG	1255									
Db	736	CTTGCCCCAGAGAAAG	751									
RESULT 14												
LOCUS	BE777259		650 bp	mRNA	linear	EST 20-SEP-2000						
DEFINITION	MY-26-C-10	Pinfestansmy	Phytophthora infestans	cdNA, mRNA	sequence.							
ACCESSION	BE777259											
VERSION	BE777259.1		GI:10230914									
KEYWORDS	EST.											
SOURCE	Phytophthora infestans		(potato late blight agent)									
ORGANISM	Phytophthora infestans		Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;									
REFERENCE	1	(bases 1 to 650)										
AUTHORS	Kamoun, S., Haberer, P. T., Sobral, B. W. S., Nuss, D. and Govers, F.											
TITLE	Initial assessment of gene diversity for the oomycete pathogen											
JOURNAL	Phytophthora infestans based on expressed sequences											
MEDLINE	Fungal Genet. Biol. 28 (2), 94-106 (1999)											
PUBMED	20056376											
COMMENT	10587472											
	Contact: Govers, F											
	Laboratory of Phytopathology											
	Wageningen University											
	Blumenhaven 9, P.O. Box 8025, 6700 EE, Wageningen, The Netherlands											
	Tel: 31 317 483 136											
	Fax: 31 317 483 412											
	Email: Francine.Govers@medew.fyto.wau.nl.											
FEATURES												
source	location/Qualifiers											
	1..650											
	/organism="Phytophthora infestans"											
	/mol_type="mRNA"											
	/strain="DDR7602, A1 mating type"											
	/db_xref="taxon:4787"											
	/dev_stage="4-week old vegetative, non-sporulating											
	mycelium in synthetic medium"											
	/lab_host="E. coli, strain DH5-alpha"											
	/clone_lib="Pinfestansmy"											
	/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; Total											
	RNA was isolated from mycelium of P. infestans DDR7602											
	cultured for 4 weeks in synthetic medium. EST clones were											
	named by their position in the microtiter plate, preceded											
	by the prefix MY (for mycelial) and the successive number											
	of the microtiter plate (e.g. MY-06-A-04)."											
ORIGIN												
	Query Match											
	Best Local Similarity	58.1%										
	Matches	319;	Conservative	0;	Mismatches	224;	Indels	6;	Gaps	1;		</

[illegible]

Db 19 ACGGCAACGATGATACAGCAAAACGACTTCTATCATCGCTACAGAGAGATCTGAAGC 78  
QY 264 TCTTGACCAAGTACGGGCAAAAGGCTTACCGCTTCTCCTTGCGTGGTCGGATCATTC 323  
Db 79 TGGTACCGACATGAATGAGACGCTTCCGCTTCTCCCTTGCAATGAGACAGGATCCTTC 138  
QY 324 CCTCGGCGGAGGCTGATCCCGTCAACGAGAGGAAATTGAGTTTACAGCAAACTGA 383  
Db 139 CAATGGACCATGCGCGAGGAAATCAAAAGCAGAGTGAATTTTACAAACAGCCTCA 198  
QY 384 TTGACGCCCTGTGATGAGCGGGGTATCAACGCTTGGGTGACTTTGTACCACTGGGATCTGC 443  
Db 199 TCGACGAGGTTTGTGCTAGAGGGCTGATGCTTTCATCAATGTTCCACTTCGACACCC 258  
QY 444 CTCAGGCGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGCTGACT 503  
Db 259 CCCAGGCCCTCGAGGACAAATACGAAAGCTTCTTGAGC--GACACATTCGTAAAGACT 315  
QY 504 TTGAGCGGTATGCGAGGTTGTCTTGAACGTTTGGGGAACGATCCAGAACTGATCA 563  
Db 316 ACGTGAGATACGGGAGCTGTGTTCAAGCTGTTGGCGACCGGGTGAAATTTGACCA 375  
QY 564 CCATCAACGAMCCTTGATTCAGGCCATCTATGATATGCCACCGGACAGACGCCCCGG 623  
Db 376 CTTTCAACGAGCCCATGATGTTCTGCGCTTCGGCTACGGCACCGGACCTTGGCCCCGG 435  
QY 624 GCAGAGC-----AGCATTAACAAGCACTCCACGAGGGCAACACTGCCACTGAGCCGT 677  
Db 436 GCGGCTGTGCGCCGTACGTCTCCAAAGGCTGCGGCGCCGGGACTCTCTCCAGGAGCCTT 495  
QY 678 GGCTCGCTGAAAGGCCCAATCATAGACCAATGCCCGCGCGGTGCGCTTACAGCAGGG 737  
Db 496 ACATCGCGGCGACACACTCTCTCATGCGCCACGCGGAGCGGTGACCTGTATCCGACCA 555  
QY 738 ACTTTCGCCCTTCGCAAAAGGCGCAATCGGCAATCTCGCTCAAGCGGACT 788  
Db 556 GTACCAAGCGGCGCAGCGGGGCCAGATCGGCAATGTCAGATATCCCACT 606

Search completed: March 28, 2004, 23:31:35  
Job time : 5351.5 secs

## SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

(without alignments)  
10962.209 Million cell updates/sec

Perfect score: 1991

Sequence: 1 agccaggtcgcgaaccagcag.....ttttaaaaaaaaaa 1991

Scoring cable: IDENTITY\_NUC

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum	DB	seq	length:	0
2000000000				

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database :

GenEmb1: \*

2: gub\_hctg:

4: gbd\_om: \*

```

5:  gub_pat:
6:  gub_pat:

```

\* : rd-qb : B  
 \* : rd-qb : B

```

gb-rc:
11:

```

12: gb\_by:  
13: gb\_un:

gd\_vt: em\_ba:

em\_hum

19: em\_mu:

21: em\_or;

23: em\_pat

25: em\_pl:

227: em\_atc8

29: em\_vi:

31: em\_htg

33: em\_ncg  
34: em\_ncg

em\_hfq

em\_by: 38:

40: em\_htg

Pred. No. is the number of results predicted by chance to have a

Query		Match		Length		DB		ID		Description	
No.	Score										
1	1454.2	73.0	1455	8	AY281377	AY281377	Hypocrea				
2	481.8	24.2	1900	8	AB003109	AB003109	Humicola				
3	452.2	22.7	1695	8	AY081764	AY081764	Hypocrea				
4	397.6	20.0	2171	8	AB003110	AB003110	Hypocrea				
5	373.4	18.8	1590	8	AY343988	AY343988	Trichoderma				
6	278.2	14.0	1828	8	AK119546	AK119546	Oryza sat				
7	278.2	14.0	2032	8	AY129294	AY129294	Oryza sat				
8	253.2	12.7	1968	8	AF268911	AF268911	Aspergillus				
9	248.6	12.5	2045	8	AK067001	AK067001	Oryza sat				
10	247.8	12.4	2068	8	AF439322	AF439322	Talaromyces				
11	246	12.4	1404	6	AX654334	AX654334	Sequence				
12	221.6	11.1	2148	8	AK067231	AK067231	Oryza sat				
13	221.6	11.1	2148	8	AK102030	AK102030	Oryza sat				
14	215.8	10.8	1811	8	AK066710	AK066710	Oryza sat				
15	214.6	10.8	1593	6	AX652977	AX652977	Sequence				
16	213	10.7	1797	8	AY056828	AY056828	Oryza sat				
17	212.2	10.7	302007	1	CSA939132	AY939132	Streptomyces				
18	210.4	10.6	1689	8	CGA266G	DB3117	Costus spec				
19	209	10.5	1665	8	AK104707	AK104707	Oryza sat				
20	205.2	10.3	1792	8	AK105908	AK105908	Oryza sat				
21	203.6	10.2	1792	8	AK098938	AK098938	Oryza sat				
22	202	10.1	1803	8	AK059210	AK059210	Oryza sat				
23	201.8	10.1	1857	8	AK103027	AK103027	Oryza sat				
24	200.2	10.1	1917	8	AK105026	AK105026	Oryza sat				
25	198.6	10.0	1894	8	AK100165	AK100165	Oryza sat				
26	198.6	10.0	1909	8	OSU78047	U28047	Oryza sativ				
27	196.4	9.9	1842	8	AK071372	AK071372	Oryza sat				
28	194.8	9.8	1824	1	CFGIJUCBET	M94865	Cellulomonas				
29	191.8	9.6	298450	1	SC039107	AL939107	Streptomyces				
30	191	9.6	2081	1	SSB0310E	Z29625	Streptomyces				
31	186.2	9.4	5996	1	AF086819	AF086819	Thermobif				
32	183.2	9.2	1796	6	AK071058	AK071058	Oryza sat				
33	182.8	9.2	1744	6	AX412252	AX412252	Sequence				
34	182.8	9.2	1734	6	AX506587	AX506587	Sequence				
35	182.8	9.2	1744	6	AX651286	AX651286	Sequence				
36	182.8	9.2	1881	8	AF159376	AF159376	Arabidops				
37	180	9.0	6219	4	OCFLP	X07995	Rabbit mRNA				
38	178.6	9.0	1533	6	AX653834	AX653834	Sequence				
39	178.6	9.0	6183	4	OCRLPH2	Z27167	O. cuniculus				
40	178.2	9.0	1562	8	AY142610	AY142610	Arabidops				
41	178.2	9.0	2016	6	AX454762	AX454762	Sequence				
42	178.2	9.0	2016	6	AX464120	AX464120	Sequence				
43	178.2	9.0	2016	6	AX491240	AX491240	Sequence				
44	178.2	9.0	2016	9	AY358729	AY358729	Homo sapi				
45	178.2	9.0	2266	6	AK574385	AK574385	Sequence				

TITLE  
Olivares,H.A., Teunissen,P.J.M., Yao,J. and Ward,M.  
Transcriptional Regulation of Biomass-Degrading Enzymes in the  
Filamentous Fungus *Trichoderma reesei*  
JOURNAL  
J. Biol. Chem. (2003) In press  
REFERENCE  
2 (bases 1 to 1455)  
AUTHORS  
Foreman,P.K., Brown,D., Dankmeyer,L., Dean,R., Diemer,S.,  
Dunn-Coleman,N.S., Goedegebuer,F., Houfek,T.D., Mitchell,T.,  
Olivares,H.A., Teunissen,P.J.M., Yao,J. and Ward,M.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (23-APR-2003) Genencor Intl., 925 Page Mill Road, Palo  
Alto, CA 94304, USA  
FEATURES  
source  
location/Qualifiers  
1..1455  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="Omega"  
/db\_xref="taxon:51453"  
/note="anamorphen: Trichoderma reesei"  
1..1455  
/gene="cell1b"  
1..1455  
/gene="cell1b"  
/note="putative beta-glucosidase"  
/codon\_start=1  
/product="Cell1b"  
/protein\_id="AAP57758.1"  
/db\_xref="GI:31747170"  
/translation="MPESTLPPNDFEMGFATAAYOISGAVEKGRPSIMPTGHLSP  
SRINGANGDVACDHYRDEADPELITTKGAAVAPSPISMSKIIPLGRLDPVNEGRLE  
FYSKLDLALRGITPPWTLTHMDLPOLHRYGMLAVEVDLFEYATLCEEREG  
DRVONWITNKPMIOALIGYATGSIAPERSINSHTEGNATPEMLAKQIMSHAR  
AAVAYSRRFRPSQKQIGISLNGDIYEPEDNEPRDKAEARMEFHLGMPANPILFK  
KDYPSMKQLEERLPAITPADFAILNAGETDFGMYVYAEQFARHLDPEPDIYD  
AIHHEOMKDGSPVSEBSGLAMLRSCPMFKHLARYGLKGYLITENCCPGRGE  
NMTCBEAVNDEPFRIRYFSDHLSISKATIDGVVVKGFAMALLDNEMSGYPRRG  
VTFTDYTLTKRTPKKSALVLMKMPAARQVRVVA"

ORIGIN  
Query Match 73.0%; Score 1454.2; DB 8; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 8.8e-285;  
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCCGATGCTTACCTCTGCTCCCAACGACTTTGATGGGCTTGGCAACGGCCGCTTAC 121  
DB 1 ATGCCCGATGCTTACCTCTGCTCCCAACGACTTTGATGGGCTTGGCAACGGCCGCTTAC 60  
QY 122 CAGATCGAAGGCGCGCTCAAGAAAGGTGGCGCGCGCGCGCTCATCTGGGACACGTACTGC 181  
DB 61 CAGATCGAAGGCGCGCTCAAGAAAGGTGGCGCGCGCGCGCTCATCTGGGACACGTACTGC 120  
QY 182 CACTGGAAGCATGCGCAGCAACGCGCGCAACGCGCGATGTGGCTTGGATCACTACAC 241  
DB 121 CACTGGAAGCATGCGCAGCAACGCGCGCAACGCGCGATGTGGCTTGGATCACTACAC 180  
QY 242 CGCTACGATGAGACTTGTATCTCTTACCAAGTACGCGCGCAAGGCGCTACCGCTTCTCC 301  
DB 181 CGCTACGATGAGACTTGTATCTCTTACCAAGTACGCGCGCAAGGCGCTACCGCTTCTCC 240  
QY 302 TTGTCTGTGTGCGCGATCATTCCTCGCGCGCGAGCTGATCCCGTCAACGAGAGGGA 361  
DB 241 TTGTCTGTGTGCGCGATCATTCCTCGCGCGCGAGCTGATCCCGTCAACGAGAGGGA 300  
QY 362 ATTGATGTTTACGACAACTGATGACGCGCTGTTGAGCGCGGATACGCGCTTGGGTG 421  
DB 301 ATTGATGTTTACGACAACTGATGACGCGCTGTTGAGCGCGGATACGCGCTTGGGTG 360  
QY 422 ACTTTGATACCATGGGATCTGGCTCAGCGCGCTTACCATGCTATGAGAGCTGGCTAAC 481  
DB 361 ACTTTGATACCATGGGATCTGGCTCAGCGCGCTTACCATGCTATGAGAGCTGGCTAAC 420  
QY 482 GTGGAAGAGGTCCAGTGAATTTGAGCGGATCGAGAGTGTGCTTGAACGTTTGGG 541  
DB 421 GTGGAAGAGGTCCAGTGAATTTGAGCGGATCGAGAGTGTGCTTGAACGTTTGGG 480

QY 542 GACCGAGTCCAGAACTGGATATCACTATCAAGAACCTTGGATTTAGGCGCATCTATGATAT 601  
DB 481 GACCGAGTCCAGAACTGGATATCACTATCAAGAACCTTGGATTTAGGCGCATCTATGATAT 540  
QY 602 GCCACCGGACAGCAACGCGCCGGGAGAGAGCATTTAACAAGCACTCCACCGAGGCAAC 661  
DB 541 GCCACCGGACAGCAACGCGCCGGGAGAGAGCATTTAACAAGCACTCCACCGAGGCAAC 600  
QY 662 ACTGCCACTGAGCCGTGGCTTCTGTGGAAGGCCCAATCATGAGCCATGCGCGCGGTG 721  
DB 601 ACTGCCACTGAGCCGTGGCTTCTGTGGAAGGCCCAATCATGAGCCATGCGCGCGGTG 660  
QY 722 GCGCTTACAGCAGGAGCTTTGCGCCCTGCGCAAAAAGGCCAAGATCCGCATCTGCTCAAC 781  
DB 661 GCGCTTACAGCAGGAGCTTTGCGCCCTGCGCAAAAAGGCCAAGATCCGCATCTGCTCAAC 720  
QY 782 GCGCATCTATGAGCCCTGGGAGCAGCAATGAGCCTCGGACAAAGAGGCTGTGAGCGA 841  
DB 721 GCGCATCTATGAGCCCTGGGAGCAGCAATGAGCCTCGGACAAAGAGGCTGTGAGCGA 780  
QY 842 CGGATGAATTTCACTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAAGCATATCCA 901  
DB 781 CGGATGAATTTCACTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAAGCATATCCA 840  
QY 902 GAGAGCATGAAGAAGCAGCTGGGGGAGAGGCTTCCAGCCCTCACTCCCGCGGACTTTGCC 961  
DB 841 GAGAGCATGAAGAAGCAGCTGGGGGAGAGGCTTCCAGCCCTCACTCCCGCGGACTTTGCC 900  
QY 962 ATCTCAATGCGGAGAGACCGAATCTTCAAGGATGAATTTACTACATCTCCAGTTGCGG 1021  
DB 901 ATCTCAATGCGGAGAGACCGAATCTTCAAGGATGAATTTACTACATCTCCAGTTGCGG 960  
QY 1022 CGGCACCTTAAAGGTCCTCCGCTCCCGGAGAGGATATCTGCGCGCCATTCATGAGCACAG 1081  
DB 961 CGGCACCTTAAAGGTCCTCCGCTCCCGGAGAGGATATCTGCGCGCCATTCATGAGCACAG 1020  
QY 1082 GAGAAATTAAGACCGGACGCGCCGTTGGCGAGAGAGGCGCTCGCTGGGCTGCGCTCTGC 1141  
DB 1021 GAGAAATTAAGACCGGACGCGCCGTTGGCGAGAGAGGCGCTCGCTGGGCTGCGCTCTGC 1080  
QY 1142 CCGGACATGTTCCGGAAGCATCTTGGCCGAGTGAAGGCTGTACCGGCAAGCCCATCTAC 1201  
DB 1081 CCGGACATGTTCCGGAAGCATCTTGGCCGAGTGAAGGCTGTACCGGCAAGCCCATCTAC 1140  
QY 1202 ATCAACGAGAAGCGGATGCGCGCTGCTGAGAGGAGAAATATGACCTGCGAGAGGCGCTG 1261  
DB 1141 ATCAACGAGAAGCGGATGCGCGCTGCTGAGAGGAGAAATATGACCTGCGAGAGGCGCTG 1200  
QY 1262 AACGACCCCTTCCGATCCGCTTACTTTGACTGCACTTGAATTCGAAAGGCAATT 1321  
DB 1201 AACGACCCCTTCCGATCCGCTTACTTTGACTGCACTTGAATTCGAAAGGCAATT 1260  
QY 1322 ACCCAGAGCGGCGTCTGCTCAAGGGGTAATTTGCTGGGCGTGTGCTGATTAATTGGAA 1381  
DB 1261 ACCCAGAGCGGCGTCTGCTCAAGGGGTAATTTGCTGGGCGTGTGCTGATTAATTGGAA 1320  
QY 1382 TTGTCAGATGGGCTACGGAACCGCATTCGCGGTCAGTTTACAGATCAACCAACCTCAAG 1441  
DB 1321 TTGTCAGATGGGCTACGGAACCGCATTCGCGGTCAGTTTACAGATCAACCAACCTCAAG 1380  
QY 1442 CGGACGCCCAAGAAAGTCTGCTGCTCAAGAGCATTTTGGCGCCCGGAGAGAGGTT 1501  
DB 1381 CGGACGCCCAAGAAAGTCTGCTGCTCAAGAGCATTTTGGCGCCCGGAGAGAGGTT 1440  
QY 1502 AAAATGGCGGATTA 1516  
DB 1441 AAAATGGCGGATTA 1455

RESULT 2  
AB003109  
LOCUS  
AB003109 1900 bp DNA linear PLN 25-DEC-2002

**DEFINITION** Humicola grisea var. thermoidea bgl4 gene for beta-glucosidase, complete cds.

**ACCESSION** AB003109

**VERSION** AB003109.1 GI:4249559

**KEYWORDS** beta-glucosidase.

**SOURCE** Humicola grisea var. thermoidea

**ORGANISM** Humicola grisea var. thermoidea  
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.

**REFERENCE** 1  
Takahima, S., Nakamura, A., Hidaka, M., Maesaki, H. and Uozumi, T.  
Molecular cloning and expression of the novel fungal  
beta-glucosidase genes from Humicola grisea and Trichoderma reesei  
J. Biochem. (1999) In press

**JOURNAL** 2 (bases 1 to 1900)

**AUTHORS** Uozumi, T.

**TITLE** Direct Submission

**REFERENCE** Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo,  
Department of Biotechnology, Faculty of Agriculture, 1-1-1 Yayoi,  
Bunkyo-ku, Tokyo 113, Japan (E-mail: uozumi@mbb.ri.a.u-tokyo.ac.jp,  
Tel: 03-5684-0387, Fax: 03-5684-0387)

**FEATURES**  
source  
1..1900  
/organism="Humicola grisea var. thermoidea"  
/mol\_type="genomic DNA"  
/sub\_species="IFO9854"  
/db\_xref="taxon:5528"  
join(289..332,437..1823)  
/gene="bgl4"  
join(289..332,437..1823)  
/gene="bgl4"  
/codon\_start=1  
/product="beta-glucosidase"  
/protein\_id="BA074958.1"  
/db\_xref="GI:4249560"  
/translation="MSLPDPFKWGFATAYQIEGVSNEDEGFSIMDTFCALDKIAD  
GSSGAAVACSYYKTKEDIALKELGANSYFSSISWSRIFPLGRNPINOKGIDHYK  
FVDDLLEAGITPEITLPHMDLPDLKRYGFLNKEFADFNVAIRIMKAIKPKCK  
WTFENPMSALIGYNTGYFAPRGTSDRSRSPVDSAREPMWIGHNILLAHARAVAY  
RDEPKPTGGEGITLNGDATLPMDPEPDADICARKEFALISWPADEYTPKGYDS  
MRQQLDRLEPFPPEVALYKGSNDPFGMHTYANTYKHTGTPPEDDFLGNLETFY  
NKYDDICIGPSTOSFWLRPHAGFRDLNMLSKRYGPKYVTGNSLKGENDMPLEQ  
VLEDPRVKYFNDYVRAAAMAAVAEDGCVNAGYLAWSLDNFEWABGEYTRFGTYVDY  
ANDQKRYPKSAKSLKPLFDSLIRKE"

**ORIGIN**

Query Match 24.2%; Score 481.8; DB 8; Length 1900;  
Best Local Similarity 61.7%; Pred. No. 2e-87;  
Matches 854; Conservative 1; Mismatches 513; Indels 15; Gaps 5;

108 CAACGGCCGCTACAGATGAAAGGCGCGTCAAGAAAGGTGGCGCGCGCCGTCATCT 167  
Db 427 CAATCAGAGCTACAGATGAGGAGCTCCGTCAGAGGATGGCGGCGCGTCATCT 486  
Qy 168 GGGACAAGTACTCCACCTGAGCCATCGGCAACCAAGCGCGCAACGGGATGTGCTT 227  
Db 487 GGGACAAGTACTCCGCGCATCCCGGCAAGATCGGCAAGCTGGGGGCGCGTGGCT 546  
Qy 228 GCGATCTACGACCGCTACGATGAGACTTTGATCTTTGACCAAGTACGGCGCAAG 287  
Db 547 GCGACTCTGACCAAGCGGACCAAGAGGACATGCGCTCTCAAGAGAGCTGGGCGCAACT 606  
Qy 288 CCTACGCTCTCTCTGTGCGTGGTGGGATCATTTCCCTCGGCGGAGGCTGGATCCCG 347  
Db 607 CGTACGCTCTCTCATCTCGTGGTGGGATCATTTCCCTCGGCGGAGGCGCAATGACCCA 666  
Qy 348 TCAACGAGAGGAGATTGAGTTTACGACAACTGATGAGCGCTGTGAGGCGGGGTA 407  
Db 667 TCAACGAGAGGAGATGACACGATGATCAAGTTTGTGCGAGACTGATGAGAGCGGCA 726  
Qy 408 TCACGCTTGGGATCTTTGATACCTGGGATCTGCTCAGGGCTTACGATGCTGATG 467  
Db 727 TCACGCTTTCATCAGCTCTTCCACTGGGACCTGCGCGAGCGCTTCGACCAAGCGTACG 786

Qy 468 GAGCTGAGCTCAACGTGGAAGAGTCCAGCTGACTTTGACGGGTATGCGAGGTGTGCT 527  
Db 787 GCGGCTTCTCTCAACAGAGAGAGGATTCGCGCGCATTTGAGAACTAGCCCGCATATGT 846  
Qy 528 TTGAAGCTTTGGGAGCCGAGTCCAGACTGTATCAACATCAAGAACCTGTGATTCAG 587  
Db 847 TCAAGCCATCCCAAGT---CAAGCAGTGTATCACTTCAACGAGACCTGTGTGCTCCG 903  
Qy 588 CCATCTATGATATGACACCGGAGGAGAGCCCGGGGAGAGACAGATTAACAGACT 647  
Db 904 CCATCTCTGCTCAACACAGGAGTACTTGGCCCCGACACACTCGAGCCGAGCAAGT 963  
Qy 648 CCAACGAGGCAACACTGCACTGAGCCGCTGCTGCTGGAAGAGCCAGATCATGAGCC 707  
Db 964 CCCCCGTGGGAGACAGCGCCCGGAGCCCTGATTTGTCCGCCAACAATCTCTATGCCC 1023  
Qy 708 ATCCCGCGCGCTGCGCTGTACAGAGGACTTTGCCCCCTGCAAAAGGCGCAATCG 767  
Db 1024 ACCCGCGCGCTCAAGGCTTACCGGAGGACTTCAAGCCCAACCCAGGCGCGAGATCG 1083  
Qy 768 GCATCTGCTCAACGGGCACTACTATGAGCCCTGGGAGACAGCAATGAGCTCGGACAAAG 827  
Db 1084 GCATCAACCTGAACGGCGAGCGCACCTCTGGAGACCCGAGACCCGCGGACATTTG 1143  
Qy 828 AGGCTGCTGAGCAGCGATGGAATTTACATTTGGCTGTGTTGCCAATCCCATCTTCTGA 887  
Db 1144 AGGCTGCGACCGCAAGATGAGTTGCGCATCTCGTGTGTTGCGGACCCCATTTACTTG 1203  
Qy 888 AGAAGGATATTCAGAGAGATGAAAGACAGCTGGGCGAGAGGCTTTCAGCCCTCACTC 947  
Db 1204 GCAAGTAC---CCGACTCGATGCGCAGACAGCTGGGCGACCGCTGCCGAGTTCACTC 1260  
Qy 948 CCGCGGACTTTGCACTCTTCAATGCCGAGAGACGCACTTTCTACGAGATTAATTAACA 1007  
Db 1261 CCAAGAGGCTGCGCTGTGCA---GGGCTCAAGCACTTTACGCGATGAACACACTACA 1317  
Qy 1008 CATCCAGTTTCGGCGCCCACTTGAAGAGTCCGCTCCCGAGAGCACTATCTCGCGCCA 1067  
Db 1318 CGGCAACATCAACAGCACAAGACGGGCTGCGCGCCGAGAGCACTTCTCGGCAAC 1377  
Qy 1068 TCCATGAGCACGAGAAATTAAGAGCGGACGCCCTGTGGCGAGAGAGCGGCTCGCT 1127  
Db 1378 TCGAGAGCTCTTTTCAACAAAGTACGGGACTGATCGGCGCGGAGACCACTCTTCT 1437  
Qy 1128 GGTGCGCTCTGCGCGGACATGTTCCGGAAGATTCGCGCGGATGAGGCTGTACG 1187  
Db 1438 GGTGCGCGCGGACGCGCCAGGCTTCCGAGACTGTCTCAACTGCTCAAGCGGCTACG 1497  
Qy 1188 GCAAGCCC---ATCTACATCAACGAGAGAGAGATGCGCTGCGCTGAGAGAGAACATGA 1244  
Db 1498 GCTACCCCAAGATCTACGAGACCGAAGACGCACTCGCTCAAGGCGAGAACGACATGC 1557  
Qy 1245 CGTGGAGAGGCGGCTCAACGACCCCTTCGATTCGAGTACTTTGATCGCATTTGACT 1304  
Db 1558 CCTCGAGAGGCTCTGAGAGACGACTTCGCGGTCAAGTACTTCAACGACTAGTGGCG 1617  
Qy 1305 CGATTTTCAAGGCTATTAACGAGAGCGGCTGTGCTCAAGGGGTACTTTGCTGGGCGT 1364  
Db 1618 CATGGCGGCGCGGCTGCGGAGAGAGGCTGCAACCTTCGCGGTTCCTGCGCTGGTCCG 1677  
Qy 1365 TGCTGATTAATTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1424  
Db 1678 TGCTGCAAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1737  
Qy 1425 ACTA---CAACACCTCAAGCGACGCGCAAGAGTCTGCTGCTGCTCAAGGACATGT 1481  
Db 1738 ACTATGCAACGACGAGAGAGGATATCCAGAAAGTGGCGCAAGAGCTCAAGCGCTGTGT 1797  
Qy 1482 TTG 1484  
Db 1798 TTG 1800

RESULT 3  
 LOCUS AY081764 1695 bp mRNA linear PLN 31-MAY-2002  
 DEFINITION Talaromyces emersonii beta-glucosidase (bg11) mRNA, complete cds.  
 ACCESSION AY081764 GI:21284364  
 VERSION AY081764.2 GI:21284364  
 KEYWORDS  
 SOURCE Talaromyces emersonii  
 ORGANISM Talaromyces emersonii  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; Talaromyces.  
 REFERENCE 1 (bases 1 to 1695)  
 MURRAY, P.G., COLLINS, C.M. and TUOHY, M.G.  
 TITLE The molecular cloning of the beta-glucosidase genes of Talaromyces emersonii  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1695)  
 MURRAY, P.G., COLLINS, C.M. and TUOHY, M.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-MAR-2002) Biochemistry, National University of Ireland, Galway, Galway, Ireland  
 3 (bases 1 to 1695)  
 REFERENCE MURRAY, P.G., COLLINS, C.M. and TUOHY, M.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Galway, Ireland  
 REMARK Sequence update by submitter  
 COMMENT On May 31, 2002 this sequence version replaced gi:19568099.  
 FEATURES  
 SOURCE  
 Location/Qualifiers  
 1..1695  
 /organism="Talaromyces emersonii"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:68825"  
 1..1695  
 /gene="bg11"  
 107..1576  
 /gene="bg11"  
 /EC\_number="3.2.1.21"  
 /note="family 1 glycosyl hydrolase"  
 /codon\_start=1  
 /product="beta-glucosidase"  
 /protein\_id="A169551.2"  
 /db\_xref="gi:21284365"  
 /translation="MTTTEPTTTTSTLPPDLMGFATSYOIEGAFDEBGRLPSTIM  
 DTFSTKPGKVDGNDVACDSYHRTGEDIILKKGAKYRFLSPRIIPLEGND  
 PINKGQFYSKFDLHAAGIEPVLFDLPELMKRYGKLNKEEFVADANYA  
 RVNENALGSKVHWITFNEPSCSVLHGNDKAPGTSRSTSPEDGPREPVIYH  
 NLVAHGTVVDIYRERFKOGEGIGITGNDMAEPMDPNRPAVEKCDKIEPAISM  
 PADPIYHKYDSNVKQGLRPLKFTFBEIAFVHGSNDFYGMNHYCENYLRNRTGEPD  
 PEDTAGNDLIMEDKSNPITGPEYOCMLRPFPLGFRKLKMLADRNPNKRIYVTENG  
 TSVKGSDDKPLEEVLNDEFRVQYRDYIGAMVAVADGVAVKAYVMAMSLDNFEMSE  
 GYRSRFGVTYVDYKNGQKRIPIKSAVLIGELFNKYRKE"

## ORIGIN

Query Match 22.7%; Score 452.2; DB 8; Length 1695;  
 Best Local Similarity 59.5%; Pred. No. 2e-81;  
 Matches 839; Conservative 1; Mismatches 559; Indels 12; Gaps 4;  
 QY 78 CTCTGCCCAAGACTTTGAATGAGGCTTCGCAAGCGCGCTTACCAAGTGAAGCGCGC 137  
 Db 147 CTCTTCCTCCGACCTTCCTCGGGGTTTCGCAATGCGACGTACCAATGCAAGGGGCTT 206  
 QY 138 TCAAGAAGAGTGGCGCGCGCCGCTTCATCTGGGACAGCTACTGCGACCTGAGGCATCGC 197  
 Db 207 TCGACGAGAGTGGCGCTGCTTCGATATGGGACACTTTCAGCAAGACGCGCGCAAGG 266  
 QY 198 GCACCAACGCGGCGCAAGCGGAGTGGCTTCGCAATGCGACCGCTACAGATGAGAGACT 257  
 Db 267 TCGAGGACGGGACCAATGGCGATGTCGTCGACCTGACCAACCGTACCGGTGAGACA 326  
 QY 258 TTGATCTCTTGACCAAGTGGCGCAAGGCTTACCGCTTCTCTTGTGTCGTCGCGGA 317  
 Db 327 TTGAGATATTGAAAGAAATAGCGGTGCCAAGATCTATTCCTTCTCTGCGCACGTA 386

QY 318 TCATTCCTTCGCGGCGAGCTGATCCGTCACAGAGAGGGAATTTGATTACAGCA 377  
 Db 387 TCATTCCTTCGCGGCGAGCAAGCAATCATATGAAAAAGGGCGTCGACTTACTTCCA 446  
 QY 378 AACTGATTGAGCCCTGTGAGGCGGGGTATCAAGCCCTTGGGTGACTTTGACACTGGG 437  
 Db 447 AGTTCTTGATGACTTACAGCGCGGCGCATTCAGCCCTTTCGTCACGCTTTTCACTGGG 506  
 QY 438 ATTCGCTCAGCGCTTACAGATCGCTATGAGAGCTGCTCAACGTGGAAGGTCAGC 497  
 Db 507 ATCTGCCGACGAGCTGATGAGCGGTACCGGGGCAATGCTCAACAGAGAGATTGTGG 566  
 QY 498 TGAATTTGAGCGGTATGCGAGGTGTGCTTTGACGTTTGGGACCGGATCCAAACT 557  
 Db 567 CCGACTAGCGCAACTATGCTCCGTGTCTTCAACCGCGCTCGGAAGTCAAGACT 626  
 QY 558 GATCAACATCAACGAMCCCTGATTCAGGCACTTATGATATGCAACCGGAGCAACG 617  
 Db 627 GATCACTTTCAGAGCCCTGTGTCAGTTCCGTGCTCGGACCAACACCGGAGACG 686  
 QY 618 CCGCGGAGAGAGAGCATTAACAGCACTTCAACGAGGCAACCTGCACTGAGCGGT 677  
 Db 687 CCGCGGCGGAGAGAGCATTCGAGCAAAAGTCCGAGGAGACGCGCACTCGCAACTT 746  
 QY 678 GGCCTGCTGGAAGGCGCCAGATCATAGAGCCCGCGCGCTGCTTACAGAGG 737  
 Db 747 GATCTGCTGCTCAACACTTCTGTGCTATGTAACGTTGAGACATCTATTCGAGAG 806  
 QY 738 ACTTTCGCCCTTCGCAAAAGGCGCAGATCGGCATCTTCGCTCAACGCGCATCTATAGC 797  
 Db 807 AGTTCAAGAGAAACAAGGCGGGGAATGGATATCAACCTCAACGCGCATGGCGGCAAC 866  
 QY 798 CTTGGACAGCATATGAGCTTCGAGCAAGAGAGCTCTGAGCGACGATGAAATTTCA 857  
 Db 867 CATGGATTCGCAAAACCAAGAGAGCTGCAACCTGCAACCGAAGATGAAATTCGCA 926  
 QY 858 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAGACTATCCAGAGAGATGAAGAG 917  
 Db 927 TCTGTGTTGTTGGGAGACCGATTTACAGGGAATAC--CGGACAGATGTTGTAAC 983  
 QY 918 AGCTGGCGAGAGGCTTTCAGCCCTCACTCCGCGGACTTTGCTCATTCATGCGGAG 977  
 Db 984 AGTTGGCGGACCGTCTGCGAAGTTTCAACCGGAAATCGGCTTCGTCGA--CGGCA 1040  
 QY 978 AGCGCACTTCAAGGCAATTAATTAACAATCCCAATTCGCGCGCCACTAGAGGTC 1037  
 Db 1041 GCAACGATTTTACGCGATGAAACACTACTCGGAAATCAATTCGCAACCGGACGCG 1100  
 QY 1038 CCGTCCCGAGAGCGACTATCTCGCGCCATCCATGAGCACAGAGAAATTAAGAGCGCA 1097  
 Db 1101 AACCAACCGGAGAGCATTTGCTGCAACTGGAACATCTGATGAGAGACAAAGACGCA 1160  
 QY 1098 GCGCCGTTGGCGAGAGAGCGGCTGCTGCTGCTGCTGCTGCGCATGTTCCGGA 1157  
 Db 1161 ACCCATTCGCGCCCGAGACCAATGCGAGTGGCTGCGCGCTGCTGCGGTTTACGGA 1220  
 QY 1158 AGCATCTCGCGCGGTGTACGCGCTGTACGGAAGCC--ATCTACATCACCGGAAG 1214  
 Db 1221 AGTTGCTGAAGTGTGCTTGCAGACCGGTACCAACCCCAAGATCTACGTGACCGGAAG 1280  
 QY 1215 GATGCCGTCCTGAGAGAGAAATGACGCTGCGAGAGGCGCTCAACGACCCCTTCC 1274  
 Db 1281 GAGCGAGCTCAAGGCGAGAGAGCAACAGCCCTGAGAGAGGTGCTCAACGATGATTCC 1340  
 QY 1275 GCATCCGTACTTTGACTGCACTTGAATCGATTTTCAAGGCTATTCGAGAGCGCG 1334  
 Db 1341 GCGTGAAGTACTACCGGGACTATATCGGCGCATGTGATCTGTGACAGGATGGGG 1400  
 QY 1335 TCGTGTCAAGGAGTACTTGTGCGGCGTTCGATTAATCTGGAATGTGATGAGTGGCT 1394  
 Db 1401 TCAAGTCAAGGCTTACATGCTTGGAGTTGTGACCACTTCAATGATGTCGAAGGCT 1460





QY 1360 GCGCTGCTCGATTAATTGGAATGCTCAGATGCTACGACCCAGATTGCGCGTCAAGTT 1419  
Db 1879 GTCGCTCATGAGCACTTTGAGTGGCGGACGCGCTACGTACGAGGTTTGCGGTTACGTA 1938  
QY 1420 CACGACT---ACACCAACCTTCAGCGCAGCCCAAGAGTCTGCTGCTCAAGGA 1476  
Db 1939 TGTGATTATGAGATGGCGCAGAGCGGTTCCCAAGAGAGCCCAAGAGCTTGAAGCC 1998  
QY 1477 CATGTTTG 1484  
Db 1999 GCTGTTTG 2006

RESULT 5  
AY343988  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AY343988 1590 bp mRNA linear PLN 13-AUG-2003  
Trichoderma viride strain AS 3.3711 beta-glucosidase 2 (bg12) mRNA,  
complete cds.  
AY343988  
AY343988.1 GI:33521681  
Trichoderma viride  
Trichoderma viride  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;  
Trichoderma.  
1 (bases 1 to 1590)  
Liu, B. D., Yang, Q., and Zhou, Q.  
Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from  
Trichoderma viride AS 3.3711  
Unpublished  
2 (bases 1 to 1590)  
Liu, B. D., Yang, Q., and Zhou, Q.  
Direct Submission  
Submitted (15-JUL-2003) Department of Life Science and Engineering,  
Harbin Institute of Technology, Harbin, Heilongjiang 150001, P. R.  
China

FEATURES  
source  
location/Qualifiers  
1..1590  
/organism="Trichoderma viride"  
/mol\_type="mRNA"  
/strain="AS 3.3711"  
/db\_xref="taxon:5547"  
1..1590  
/gene="bg12"  
192..1544  
/gene="bg12"  
/codon\_start=1  
/product="beta-glucosidase 2"  
/protein\_id="AA021384.1"  
/db\_xref="GI:33521682"  
/translation="MLPKDFQWGFATTAIYQIRGAVDQGRGSIWDTCAQPKIADG  
SSGVTAQSYNRTADIALSLKSLGAKSYRFSISWRIIPGGRGDAVNOAGIDHYKVF  
VNDLDAGITTPITLTFHMDLPEGLHQRVGLNRTPEPLDENYAVNVRALPKVRN  
ITFNPICSAIPGYSGTFAPRGOSTSPWYGNHILVAHGAVAVARDPFAAGD  
GIGIYNGPFTYPMADPADKEARELERLEPTAFADPIYAGDYPAARMKOLGRLP  
TPTPERALHGSNDPYGMNHYTSVTHRSRSPATTSATSTGCSPLSPATSAAPR  
SPPLAPVRPAATSMCGSARGTTPITYTENGSITGESDLPREKILBEDPFRKYNEY  
IRAWTVAVELDGVNKGYPAWSLMDNPEMADGYVTRFGVTVYENGGKRSRRAQRA  
"

ORIGIN  
Query Match 18 8%; Score 373.4; DB 8; Length 1590;  
Best Local Similarity 59.4%; Pred. No. 2e-65;  
Matches 802; Conservative 1; Mismatches 497; Indels 51; Gaps 8;

QY 81 TGCCCAACGACTTTGATGATGGGCTTCGCAACGCGCCCTACCAAGATCGAAGCGCGCTCA 140  
Db 196 TGCCCAAGAGCTTTGATGATGGGCTTCGCAACGCGCTTCCTACCAAGATCGAAGCGCGCTCG 255  
QY 141 AAGAAAGTGGCGCGCGCGCTTCATTTGGAGACAGCTGCTCCACCTGGAGCATGCGCA 200  
Db 256 ACCAGACGCGCGCGCGCGCTTCGAGACAGCTTCTGCGCGAGCCCGCAAGATCG 315

QY 201 CCAACGCGCGCAACGCGATGATGGCTTGGCATCACTACACCGCTACGATGAGACTTTTG 260  
Db 316 CCGAGCGCTGTGGGGGTGACCGGCTGCGACTGCTGACACCGGACCGCGAGAGACTTGG 375  
QY 261 ATCTCTTGACCAAGTACCGGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGGATCA 320  
Db 376 CGGTGCTCAAGTGTGCTGCGGCGCAAGGCTTACCGCTTCTCTGCTGCTGCGGATCA 435  
QY 321 TTCCCTCGCGCGCGCGCTGAGATCCCGTCAACGAGAGGGAATTGATTTCACGCAAC 380  
Db 436 TCCCGAGCGCGCGCGCGGATGCGCTCAACGAGGCGGATGACACTACGTCAAGT 495  
QY 381 TGAATGACCGCTGTTGATGAGCGGCTTACACGCTTGGGTGACTTTGTACCACTGGAGATC 440  
Db 496 TCGTCAACGACCTGCTGACGCGCGGATCAACGCTTTCATCACTTTCATGCGGAGC 555  
QY 441 TGCTCAAGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTGAAAGAGTCCAGCTGG 500  
Db 556 TGCCGAGGCGCTGATCAGCGGCTACGCGGCGCTGCTGAACCGCACCGAGTTCGCGCTCG 615  
QY 501 ACTTTGACCGGTATGCGAGGTGTGCTTTGAACGTTTGGGACCGAGTCCAGAACCTGGA 560  
Db 616 ACTTTGAAAATGACCGCGGTCAATGTTCAAGGCGCT--GCCAAGGTGCGCACTGGA 672  
QY 561 TCACCATCAACGACCTGATTCAGGCTATGATGATATGATGATGACCGGAGCAACGCC 620  
Db 673 TCACCTTCAACGACCGCTGCTGCTGCGGCTATCCGCGCTAGCGGCTCGGACCTTGGCCC 732  
QY 621 CGGCGAGGAGCAGCATTAACAGACCTCCACCGAGGCAACCTGCCACTGAGCGCTGGC 680  
Db 733 CGGCGCGGAGAGCA-----CCTCGAGCGCGTGA 762  
QY 681 TCGCTGGAAGGCCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 740  
Db 763 CGGTGCGCACACATCTGCTGCGCCACGCGCGCGCTGACAGGCTGATCCGAGACACT 822  
QY 741 T---TCGCCCCCTCGCAAAAGGCCAGATCGGCACTGCTCAACGCGGACTATATGAGC 797  
Db 823 TCAAGCCCGCAAGGCGGAGCGGCGAGATCGGCACTGCTCAACGCGGACTATATGAGC 882  
QY 798 CTTGGAACAGCATGATGAGCTTCGCGGACAAAGAGCTGCTGAGCGCAGATGGAATTTCACA 857  
Db 883 CTTGGAACAGCATGATGAGCTTCGCGGACAAAGAGCTGCTGAGCGCAGATGGAATTTCACA 942  
QY 858 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAGACTATCCAGAGAGCATGAAGAGC 917  
Db 943 CGGCTGTTTGGCAATCCCATCTTCTTGAAGAGACTATCCAGAGAGCATGAAGAGC 999  
QY 918 AGCTGGGAGAGGCTTCAGACCTCACTCCCGGAGACTTTGCAATCTCAATCCGAG 977  
Db 1000 AGCTGGGAGAGGCTTCAGACCTCACTCCCGGAGACTTTGCAATCTCAATCCGAG 1056  
QY 978 AGACGACTTTACGCGATGATTAATTAACATATCCAGTTGCGCGGCACTTACAGCTC 1037  
Db 1057 CCAACGACTTTTACGCGATGATTAATTAACATATCCAGTTGCGCGGCACTTACAGCTC 1116  
QY 1038 CCGTCCCGGAGAGGACTATCTCGCGCGCATCCATGAGACACGAGAGATTAAGAGCGCA 1097  
Db 1117 CCGGCTC---GACGACACCGTCCGCAACGTGACGTGCTTTTACCAACAAAGAGGCA 1172  
QY 1098 GCCCGTTTGGGAGAGAGCGGCTTCGCTGCGCTGCTGCGCGGACATGTTCCGGA 1157  
Db 1173 ACTGATGAGGCGCGGAGAG--CGACAGTCCCTTGGCCCTGTGCGCCTGCGGACATTC 1231  
QY 1158 AGCATCTGCGCGGCTGATGAGCTGTAAGCGCAAGCCATTAATCAACGAGAGCGAT 1217  
Db 1232 CTGTGTGATCAACCAAGGTAC---CTAACCGCCATTAATGATGAGAGAGCGCA 1287  
QY 1218 GCCCGTGCCTGAGAGAGAGCAATGAGTCCGAGAGAGCGGCTGAACGACCCCTTCCGCA 1277  
Db 1288 CGAGCATCAAGGCGAGAGCACTTGGCCAGAGAGAGATTTCTGGAAGATGACTTCAGGG 1347

QY	1278	CCGSGTCTTGCATCTGGACCTTGACCTGCATGTTCCAGAGCCATTACCCAGAGCGCGTC	1337
DB	1348	TCMACTACTATACGACGATACATCCGTCGATGTTACCGCCGTGAGCTGAGCGGGTCA	1407
QY	1338	TCGTCMAAGGGGTACTTTCGTCGTGGCGTTGCTCGATTACTTGGAAATGTCAGATGCTTACG	1397
DB	1408	ACGTCAAGGGGATCTTTCCTGTGTGTCGTCATGACCAACTTTGAGTGGCGGAGCGGCTACG	1467
QY	1398	GACCCAGATTGGCGCTACGTTACAGACTA	1428
DB	1468	TGACGAGGTTGGGGTTACTATGTGATTA	1498
RESULT 6			
AK119546			
LOCUS			
DEFINITION	AK119546	1828 bp	mRNA linear
ACCESSION	AK119546		
VERSION	AK119546.1	GI:37989169	
KEYWORDS	Full-CDNA; oligo capping.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE	1		
AUTHORS	The Rice Full-Length cDNA Consortium, National Institute of		
	Agronomical Sciences Rice Full-Length cDNA Project Team,		
	Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,		
	Kishimoto, N., Yazaki, T., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,		
	Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C.,		
	Ohtsuki, K., Shieniki, T., Foundation of Advancement of International		
	Science Genome Sequencing & Analysis Group, Otsu, Y., Murakami, K.,		
	Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tanoda, Y.,		
	Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,		
	Naikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikkura, J.,		
	Ikedo, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mura, J.,		
	Kusunegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,		
	Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,		
	Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,		
	Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Otsu, N., Oka, Y.,		
	Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,		
	Yoshino, M., and Hayashizaki, Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from		
JOURNAL	Japanese rice		
MEDLINE	Science 301 (5631), 376-379 (2003)		
PUBMED	22752273		
REFERENCE	12869764		
AUTHORS	2		
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,		
	Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, M.,		
	Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T.,		
	Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,		
	Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,		
	Kawagashira, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,		
	Kikuchi, T., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,		
	Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,		
	Koya, S., Kurihara, C., Kurosaki, T., Kusunegi, T., Li, C., Lu, M.,		
	Masuda, H., Matsubara, K., Matsuyama, T., Mura, J., Miyazaki, A.,		
	Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahara, T.,		
	Nakamura, M., Namiki, T., Naikawa, R., Nikkura, J., Nishi, K.,		
	Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M.,		
	Ooka, Y., Otsu, N., Ota, Y., Otsu, Y., Ryu, R., Satoh, H., Sasaki, C.,		
	Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,		
	Shibata, K., Shinagawa, A., Shiraki, T., Shieniki, T., Sogabe, Y.,		
	Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, K., Tagami, M.,		
	Tagami, Takeda, Y., Tagawa, A., Takahashi, F., Takaki-Akashita, S.,		
	Tanaka, T., Tomaru, A., Toy, T., Tanoda, Y., Ueda, M., Waki, K.,		
	Xie, Q., Yanagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,		
	Yokomizo, S., and Yoshimura, A.		
TITLE	Collection, mapping, and annotation of 28k full-length cDNA clones		
	from japonica rice		

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 1828)  
 AUTHORS Kikuchi, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of  
 Agricultural Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-833-7007)  
 This clone is one of the 32k full-length cDNA clones from japonica  
 rice.  
 URL : [http://cdna01.dna.affrc.go.jp/cDNA/NIAS\\_Rice\\_Full-length\\_cDNA\\_Project/Team\\_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,Y.,Nami,T.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hota,I.,Kojima,K.,Nishiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.,Yanamoto,M.,and Nakahama,Y.](http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_cDNA_Project/Team_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,Y.,Nami,T.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hota,I.,Kojima,K.,Nishiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.,Yanamoto,M.,and Nakahama,Y.)  
 FAS Genome Sequencing & Analysis Group: Oono, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kodayashi, M., Kodama, T., Kuroaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nakikawa, R., Nikiya, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matubara, K. and Murakami, K. Genome Exploration Research Group in Riken Genomic Sciences Centre and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hangaki, T., Hara, A., Haseizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numaaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Satazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takehachi, F., Takaki-Akahita, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yaeunishi, A. and Hayashizaki, Y. Location/Qualifiers  
 1. 1828  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultiyar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="001-207-A05"  
 ORIGIN  
 Query Match 14.0% Score 278.2; DB 8; Length 1828;  
 Best Local Similarity 54.7% Pred. No. 4.2e-46;  
 Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;  
 Oy TGCCCAAGCACTTGAATGGGCTTTCGCAAGCGCGCCTTACAGATCGAAGCGCGCTCA 140  
 Db TCCCCGAGGGTTGCTCTTGGCACCGCTCCGCCGCGTACCAAGTACGAGGAGCTGTGA 267  
 Oy 141 AAGAAGGTGGCGCGCGCCGTCATCTTGACACAGTACTGACACCTGAGCCATGCGCA 200  
 Db 268 AGGAGGACGGGAGGCGACAGCACTTGGGACACGTTGGCCACACCTTTGGAAGAATCA 327  
 Oy 201 CCAACGGCGCCAAAGCGGATGGCTTGGCATCACTACACCGCTACGATGAGACCTTGG 260  
 Db 328 CCGACTTCAGCAATGCTGATGTTTCAGTGTACGTACCAACGCTTTCAGAGGAGTATAC 387  
 Oy 261 ATCTCTTGACCAAGTACGGCGCAAGGCGTACCGCTTCTCTTGCAGTGTGCGGATCA 320  
 Db 388 AGCTCATGCGACGACATGGGATGATGCTGTTCGCTTTCGATAGCATGTGTCAAGATCT 447  
 Oy 321 TTCCCTTCGCGGAGGCTGTATCCGTCAACGAGAGGAATTAAGTTTATACGCAAC 380  
 Db 448 ACCCAATATGGG-----TTGTCAAGTCATCAAGCTGTGATCGAACCAATCAAGC 501  
 Oy 381 TGATTGACGCTGTGAGGCGGGGTATACGCTTGGGTACTTTGTATCAACTGGATC 440

Db 502 TGAATGATGACCTTCTAGCAAAAGAAATTCAGCCATATGTGACACTTACCATCTGGAGACC 561  
QY 441 TGCCCTGAGCGCTTACGATCGCTATGAGGCTGGCTCAAAGTGAAGAGGTCCAGCTGG 500  
Db 562 TCCCCGAGCGCTTGAAGAACAGTAACAAGGCTGCTTGAACAGCA---GATAGTGAAGC 618  
QY 501 ACTTTGAGCGGATGCGAGGTTGTCTTTGAACGTTTGGGAGACCGAGTCCAGAACTTGA 560  
Db 619 ATTTGCGGGGTACGCGAGACGCTGCTTCAGGAGTTTCGGGAGACAGGATGAAGCACTTGA 678  
QY 561 TCACCATCAACGACCTCGATTCAGGCACTCTATGATATGACCGGAGCAACCGCC 620  
Db 679 TCACGCTCAACGACCGGACACGCTGGCATTCAGGCTACGAGCAGGAGGCTCCAGGCCC 738  
QY 621 CGGAGCAGAGCAGC-----ATTAAACAAGACTCCACGAGGAGCAACTGCGCACTAGC 674  
Db 739 CGGAGCGGTGCTCGGTGCTGCTCACCCTCTACTCAAGGCGGAGCACTCCGAGCAGCAGC 798  
QY 675 CGTGGCTGCTGAAAGGCCCAATCATAGCCATGCCCGGCGGTGGCCCTCTAACAGCA 734  
Db 799 CTTACGCTGCTGCCACCACTTCACTCTCGCCCAAGCGCGCGCAGCACTTCAAGGA 858  
QY 735 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAACGCGACTACTATG 794  
Db 859 CAAATACAAAGCGACGAGAACCGGAGCTTGGGATAGCGTTGAGCGTGAATGTGTTGG 918  
QY 795 AGCCCTGGGACAGCAGATGAGCCTCGGACAAAGAGGCTCTGAGCGACGAGTGAATTC 854  
Db 919 AGCCGATGTCCAAACCAACGATC---GACATCGAGGCGGCAAGAGACCGCAGAGATTC 975  
QY 855 ACATTTGGTGGTTTGGCAATCCCATCTTCTTGAAGAAAGCATTCACAGAGCAGTGAAGA 914  
Db 976 AGCTAGATGTGTTGCTGATCCGTTCTTCTT---CGGAGACTACCGGCGAGCAGTGAAGG 1032  
QY 915 AGCAGCTGGGAGAGGCTTCCAGGCTTCACTCCCGCGAGATTTGCCATCTCAATGCGG 974  
Db 1033 CGAGGCTGGGGAGAGGCTGCGCAGAGTTCAAGCGGATGAGGCGCGCTGCTCAAG---G 1089  
QY 975 GAGAGACCGACTTCTTACGGCATGAATTAACAATCCCATGTCGGCGGCACTTAAGAG 1034  
Db 1090 GGGCGCTGATTTCTCGGCAATTAACCACTACACCACTACTACACGAGGACCAACA 1149  
QY 1035 GTCCGCTCCCGAGA-----CGGACTATCTCGGCGCATTCACATGAGCAGAGAGAT-- 1087  
Db 1150 CCAACATATGCGGACATTTGCTCAACAACACTTGGCAGACACCGGACACGTCAGCTCC 1209  
QY 1088 -----AAGACGCGACGCCCTGTGGCGAGAGAGCGGCTGCTGGCTGGCTCTGCC 1142  
Db 1210 CATTCAAGATGGGAAGCCAATTGGAGATAGGGCAAAATTCGATATGGCTGTACATTGTGC 1269  
QY 1143 CGGACATGTTCCGGAAGCATCTCGCCCGGCTGTAAGGCTGT---ACGGCAAGCCCATCT 1199  
Db 1270 CCGGAGGATGAGAGGCTGATGAATATGTCAAGAAAGTAACAACGCCACCAAGTGT 1329  
QY 1200 ACATCACCGAGAACGATGCGCCGCTCGAGAGAGAACATGACGTGCGAGAGGCGC 1259  
Db 1330 ACATCACTGAAGAACGAGATGATGAAGCAACACCGTTCACTTTCATTAAGAGCGCCC 1389  
QY 1260 TCACGAGCCCTTCCGATCCGTACTTTGACTCGCATCTTGAATTCGATTTCAAGGCCA 1319  
Db 1390 TCACGAGACGCAAGAGATCAATATCAACAATGACTACTCAACATCTGAGTGGCTTCCA 1449  
QY 1320 TTACCCAGAGACGGCGTCTGCTCAAGGGGTACTTTGCGTGGGCGTGTGCTGATACTTGG 1379  
Db 1450 TCACGAGAGACGGGTGCGAGCTGAGTACCTGCGTGTCTGCTGAGCAACTGGG 1509  
QY 1380 AATGCTGAGATGCTACGAGACCAATTCGCGGTCAAGTTCAACAAGTCA---CCACC 1436  
Db 1510 AATGGGCGGCGGATATCTCTGAGATTCGGGCTCTACTTGTGAGCTTCAAGAGATAACC 1569  
QY 1437 TCAAGCGACGCGCAAGAACTGCTGCTGCTCTCAAGACATGTTTGGGCGCCG 1491  
Db 1570 TCACGAGATACCCCAAGAACTCGGTGCAAGGCTCTCTGAAAGACCTG 1624

RESULT 7  
AY129294 2032 bp mRNA linear PLN 03-SEP-2002  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) beta-glucosidase mRNA,  
complete cds.  
AY129294  
ACCESSION  
VERSION  
AY129294.1 GI:22658439  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE  
1 (bases 1 to 2032)  
Akiyama,T., Ogasawara,R., Ketudat-Cairns,J.R., Swast,J. and Ezen,A.  
Predominant expression of beta-glucosidase in germinating rice  
seeds  
JOURNAL  
2 (bases 1 to 2032)  
Akiyama,T., Ogasawara,R., Ketudat-Cairns,J.R., Swast,J. and Ezen,A.  
Direct Submission  
Submitted (05-JUL-2002) Low Temperature Science, National Hokkaido  
Agricultural Research Center, 1 Hitsujigaoka, Toyohira-ku, Sapporo,  
Hokkaido 062-8555, Japan  
LOCATION/Qualifiers  
1..2032  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Yukihikari"  
/db\_xref="taxon:39947"  
142..1707  
/codon\_start=1  
/product="beta-glucosidase"  
/protein\_id="AA01354.1"  
/db\_xref="GI:22658440"  
/translation="MGRIKSSSGRCSTABLEAVAVLVVFGVASSLSRGCTAQQSGG  
LTRGSPGPEGVFGTASAAVYEGAVVEDGQVITWTFATPKITDFSNADVADVY  
HRPEEDIPQALMDMGDAVRSIAMSRTYVQGVQVNOAGIDHKLITDALLAGIOPY  
VTIYHMDLPQALDPRKXGMDROIUDPAFVATCFRPEPDRIYHMTLNEPHTVALO  
GYDAGLPGRCSVLHLTYKANSRGTEPYVAHHHTLAANAASIRTYKATONCO  
LGIAFDVMEPEPMSNTTIDLEAKRAQEFQMGFADPFPGDYPATRAVGERLPRF  
TADAEAVKGLDFVGINHYTYTTRNNNTNIGTLNLTLDLTGIVSLPKNGKPIG  
DRANSILYIVPRGMSIMYKERVNSPVIYTEGMDSNPFISIKDALDSKRI  
KYHNDVLTNLAASIKEDGCDVRGYFAWSLIDNEMWAGYSRGLYFVDYKMLKRYP  
KNSVQWPKALKTI"

ORIGIN  
Query Match 14.0%; Score 278.2; DB 8; Length 2032;  
Best Local Similarity 54.7%; Pred. No. 4,2e-46;  
Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;  
QY 81 TGCCCAACGACTTGAATGGGGCTTCCCAACGCGCGCTCAACATGGAAGGCGCCGCA 140  
Db 290 TCCCGAGGGGTTGCTTCCGACCGCTCTCGCGGTACAGTACGAGGAGCTGTGA 349  
QY 141 AAGAAGTGGCGCGGCCCGCTCATCTGGGACAGTACTGCGCACTGGAGCCATCGCGCA 200  
Db 350 AGGAGAGCGGAGAGGGCAGACATCTGGGACAGTTCGGGCAACCTTTGGAAGAATCA 409  
QY 201 CCAACGGGCGCAACGGCGATGTCCTTGGCATCTACACACCGCTAGATAGGACTTGG 260  
Db 410 CCGACTGAGCAATGCTGATGTTGAGTGTACATCAACCGTTTCGAGAGGATATAC 469  
QY 261 ATCTTTGACCAAGTACGAGCCCAAGAGCTACCGCTCTCTGTGCGTGTGCGGATCA 320  
Db 470 AGCTCATGACAGACATGGGAGTGAATGCTTCCGTTTCCATAGCATGATCAAGAACT 529  
QY 321 TTCCCTCGCGGAGCGCTGATCCGCTCAACGAGAGGGAATGATTTTACAGCAAC 380  
Db 530 ACCCAATGTG-----TTGTCAAGTCAATCAAGCTGATGACCACTAACAAGC 583

QY 381 TGAATGACGCGCTTTGAGCGGGGATATCAAGCCTTGGTACTTTGTATCACTGGGATC 440  
 DB 584 TGAATGACGCGCTTTGAGCGGGGATATCAAGCCTTGGTACTTTGTATCACTGGGATC 643  
 QY 441 TGGCTCAGCGGCTTACGATCGCTATGAGAGCTGGCTCAACGTTGAAGAGAGTCCAGCTGG 500  
 DB 644 TCCCGCGGCGCTTGAAGAGAGTCAAGGCTGGCTTGAAGAGGCA---CATAGTGAAGC 700  
 QY 501 ACTTTGAGCGGATGAGAGGCTTGGCTTTGAAGCTTTGGGAGCCGAGTCCAGAACTGGA 560  
 DB 701 ATTTCGGGCTGACGAGAGAGCTTCAAGGAGTTCGGGAGAGGAGTGAAGAGACTGGA 760  
 QY 561 TCACATCAACAGAGCCCTGATTCAGGCACTTATGATATGACCGGAGAGAGAGAGAGCC 620  
 DB 761 TCACGCTCAACAGAGCCGAGAGAGTTCAGGCACTTCAAGAGAGAGAGAGAGAGAGCC 820  
 QY 621 CCGGCGAGAGAGAGC---ATTAAACAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGC 674  
 DB 821 CCGGCGAGAGAGAGC---ATTAAACAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGC 880  
 QY 675 CGTGGCTGCGTGAAGC 724  
 DB 881 CTTACGCTGCGGAGC 940  
 QY 735 GGGAGCTTCCGCGCTCGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 794  
 DB 941 CAAATATCAAGC 1000  
 QY 795 AGGCTTGGAGC 854  
 DB 1001 AGGCTTGGAGC 1057  
 QY 855 ACATTTGGCTGCTTGGCAATCCATCTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGC 914  
 DB 1058 AGCTAGAGATGTTGCTGATCCGCTTCTTCTT---CGCGAGCTAACCGGAGAGAGAGAGC 1114  
 QY 915 AGCAGCTGAGC 974  
 DB 1115 CGAGC 1171  
 QY 975 GAGC 1034  
 DB 1172 GGGCGCTGATTTGCTGAGC 1231  
 QY 1035 GTCCCGTCCCGAGC 1087  
 DB 1232 CCAACATGATGAGC 1291  
 QY 1088 ---AAGC 1142  
 DB 1292 CATTTCAAGATGAGC 1351  
 QY 1143 CGGAGATGTTCCGAGC 1199  
 DB 1352 CCGAGC 1411  
 QY 1200 ACATTCACGAGC 1259  
 DB 1412 ACATTCACGAGC 1471  
 QY 1260 TCACAGC 1319  
 DB 1472 TCACAGC 1531  
 QY 1320 TTAACAGC 1379  
 DB 1532 TCACAGC 1591  
 QY 1380 AATGCTCAGATGCTAGC 1436  
 DB 1592 AGTGGGAGC 1651  
 QY 1437 TCAAGCGAGC 1491

Db 1652 TCACAGAGATACCCCAAGAACTCGTGCTAGTGTTCAGAGCGCTCTGAGAGAGCTG 1706  
 RESULT 8  
 AF268911  
 LOCUS  
 DEFINITION  
 Aspergillus niger beta-glucosidase precursor (bg11) gene, complete cds.  
 ACCESSION  
 AF268911  
 AF268911.2 GI:8572535  
 VERSION  
 AF268911.2  
 KEYWORDS  
 Aspergillus niger  
 Aspergillus niger  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 REFERENCE  
 1 (bases 1 to 1988)  
 Al, Y.-C. and Meng, F.-M.  
 Molecular cloning and sequencing of a beta-glucosidase gene from  
 Aspergillus niger Ams11  
 Unpublished  
 JOURNAL  
 2 (bases 1 to 1988)  
 Al, Y.-C. and Meng, F.-M.  
 Direct Submision  
 Submitted (17-MAY-2000) Department of Biochemistry, Zhongshan  
 University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.  
 China  
 REFERENCE  
 3 (bases 1 to 1988)  
 Al, Y.-C. and Meng, F.-M.  
 Direct Submision  
 Submitted (19-JUN-2000) Department of Biochemistry, Zhongshan  
 University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.  
 China  
 REMARK  
 COMMENT  
 On Jun 19, 2000 this sequence version replaced gi:8248474.  
 FEATURES  
 source  
 1. 1988  
 /organism="Aspergillus niger"  
 /mol\_type="genomic DNA"  
 /strain="Ams11"  
 /db\_xref="taxon:5061"  
 <496..>1858  
 /gene="bg11"  
 join(<496..536,595..>1858)  
 /gene="bg11"  
 /product="beta-glucosidase precursor"  
 join(496..536,595..1858)  
 /gene="bg11"  
 /codon\_start=1  
 /product="beta-glucosidase precursor"  
 /protein\_id="AA074209.2"  
 /db\_xref="GI:8572535"  
 /translation="MLPKDLQWGFAPAAVQIEGAVDQDGRPSIWDTFCAQPKIADG  
 SSGVTACDSYNRPAEDIALKSLGASYSFSSSRIPBGRGDAVNAQAGIDHYKVD  
 DLIDAGITPEITLTFHMDLHORYGILNTEFLDEBENARVAFRALPKVRNNEPLC  
 SAIPGSGSPFARGQSTSEPTWGHNIIVAHGRAYKAVRDPKPSGDDQIDIVNG  
 DPTVPMADADPADKLERLFEFTAFADPIYVIGDPAEMKQIDRLPLTFPEBALVAG  
 SDFYGMNHTSYIRHRSPPASADTVGNVYLPFNKQNCIGPRTQSWLRPCAG  
 FRDLVWTSKRTGSPPIVYENGTSTKGSIDLNEKILDEDFVAKTYNETIRAMVTA  
 ELDGAVNRGVTVVDYENQKRPKKSASLKLFLDELAAA"  
 join(496..536,595)  
 /gene="bg11"  
 596..1855  
 /gene="bg11"  
 /product="beta-glucosidase"  
 ORIGIN  
 Query Match 12.7%; Score 253.2; DB 8; Length 1988;  
 Best Local Similarity 56.0%; Pred. No. 4.9e-41;  
 Matches 708; Conservative 1; Mismatches 474; Indels 81; Gaps 9;  
 QY 106 CGCAAGCGCGCTACAGATCGAAGCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGC 165

Db	583	CCCGTCCGACGAGTACCAAGATCGAGGGCGCGTGCACCAAGACGGCGCGCCCAACAT	642
OY	166	CTGGGACACGTACTGCACTCGAGGCCATGCGGACCAAGCGGCCCAACGGCCATGTCG	225
Db	643	CTGGGACACGTTCTGCGCGGACGCCCGGCAAMATGCGGACGGCTGTCGGGCGTGACGGC	702
OY	226	TTGGATCACTACCAACCGCTCGATGAGGACTTTGATCTCTTGAACCAATACGGCGCAA	285
Db	703	GTGGGACCTGTAACAACCGCACGGCGGAGACATTGCGGTGCTCAATGTCGTCGGGGCCAA	762
OY	286	GGCCTACCGCTTCTCTTGTGTGTGTGCGGATATTCCCTCGGCGGCAAGCTGATCC	345
Db	763	GAGCTACCGCTTCTCTCTCGT-----CGCGATCCCCGAGGCGCGCGCGGCGATGC	816
OY	346	CGTCAACGAGGAGGAAATTGAGTTTTCACAGAACTGATTCACGCCCTGTGTGAGCGGAG	405
Db	817	CGTCAACGAGGCGGAGCTGACCACTAGTAAATTGTCGATGACACACTGCTCGACGCCG	876
OY	406	TATCACGCTTGGGTGACTTTGTATCACTGAGGATCTGCTCAGCGCCTTCCAGTCTGCTA	465
Db	877	CATCACGCTTTCATCACTCTTCTCACTGGGACCT-----GGTGCATCAGCGGTA	927
OY	466	TGAGAGCTGAGTCAACGTGAAAGAGTCCACTGATGACTTTGAGGCGTATGAGAGTGTG	525
Db	928	CGGGGGGCTGTAACCGCACCGAGTTCCGCTGACTTTGAAAACTACGCGCCGCTCAT	987
OY	526	CTTTGAACGTTTGGGGACCGAGTCCAGAACCTGATCACTCAACGAMCCCTGATTTCA	585
Db	988	GTT-----CAGGGCGTGGCCCAAGAGTGCGCAACTGAAACGAGCGGCTGTGCTC	1035
OY	586	GGCCATCTATGAGATATGCCACCGGACAGAACGCCCGGGGACAGAGCACTTAACACGA	645
Db	1036	GGCCATCCCGGAGCTACGGCTCGGAGTTTGCCCGCGCGGACGACCA-----	1085
OY	646	CTCCACGAGGACAACTGCACTGAGCCGTGCTCGCTGAAAAGGCCAGATCATAG	705
Db	1086	-----CCTGGGACCTGTGACCGTGGGCAACATCTGTGTGC	1125
OY	706	CCATGCCCGCGCGCTGCGCTTACAGCAGGAGCTT---TGCCCTTGCAAAAGGCGCA	762
Db	1126	CCACGCGCGCGCGCTCAAGGGGTACCGGACGACTTCAAGCCCGCACGCGGACGCGCA	1185
OY	763	GATGGGACCTGCGCTCAACGGGGACTACTATGAGCCGCGGACAGCAATGAGCCTCGGGA	822
Db	1186	GATGGGACCTGCTCTCAAGGGGACTTACTACCCCTGTGGAGCGCGCGACCGGCGCGA	1245
OY	823	CAGGAGGCTGTGAGCGACGAGTGAATTTCATTGGCTGTGCTTGCAAATCCCATCTT	882
Db	1246	CAGGAG-----CGCTCGAGTCTTTCACGCGCTGAGTTGCGGATCCCATCTA	1293
OY	883	CTTGAAGAAGACTATCCAGAGAGATGAAGAAGAGCTGGGCGGAGAGGCTTCAAGCCT	942
Db	1294	CTTGGGCG---GACTACCGGCGCTCGATGGCGAAGAGCTGTGGCGACCGGCTGCACCTT	1355
OY	943	CACCTCCGCGGACTTTGGCCATCTCAATGCGCGAGAGACCGACTTCAACGCGATGAATTA	1002
Db	1351	TACGCCCGAGAGGCGCGCTCTGTCA---CGGCTCAACGACTTTTACGGCATGAACCA	1407
OY	1003	CTACACATCCAGTTTGCGCGCTCACTTGAACGATCCGCTCCCGAGACGGAATATCTGG	1062
Db	1408	CTACACGCTCAACTATCCGCGCAACGCTCGCGCGCTTCCGCGACGACACGCTGG	1467
OY	1063	CGGCATTCATGAGCACAGGAAGATTAAGAGCGGACGCCGCTTGGCGAGAGAGCGGCT	1122
Db	1468	CACGCTGACGTGCTCTTACCAACAAGAGAGGCACTGCACTCGGCCCGGAGACGCACTC	1522
OY	1123	CGCCTGAGTGCCTCTGCGCGGACATGTTCCGGAAGACTTCGCGCGGTACGCGCT	1182
Db	1528	CCCTGTGCTGCGCCCTGTCGCGCGGCTTCCGCACTTCTGTGTGACACAGAAAG	1587
OY	1183	GTAACG---CAAGCCATTAATCAACGGAAGACGATGCCGTCCTGTGAGAGAGAGA	1233
Db	1588	GTAACGCTCCCGCCATTAAGTACGTACGGAAGACGCAACGACATCAAGGGGAGAGCGA	1644

Oy		1240	CATGACGTGGAGAGAGGCCCTCAACAACACCCTTCGCATCCGGTACTTTAATTACCGCATCTT	1299
Db		1648	CTTGCCCCAACGAAGAAGTTCTTCGAAGTAGCCTTAAGSGCTAAAGTACTATAACGAGTACAT	1707
Oy		1300	GGAAGTCAATTTCCAAGAGCCATTACCCAGAGCGAGCGTGCATCGTCAAGGGGTACTTTGGCGTG	1359
Db		1708	CCGTGGCATTGGTTTACCGCCGCTGGAGCTGGAACGGGGGTCAAAGTAGGTTTTGGGGTTTACGTA	1767
Oy		1360	GGCG 1363	
Db		1768	TGTG 1771	
RESULT 9 AKO67001 LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS		AKO67001  Oryza sativa (japonica cultivar-group) insert sequence. AKO67001 AKO67001.1 GI:32977019 FLI_CDNA; CAP trapper. Oryza sativa [japonica cultivar-group] Oryza sativa [japonica cultivar-group] Eukaryota; Viridiplantae; Streptophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehharatoidae; Oryzeae; Oryza.	2045 bp mRNA linear PIN 24-JUL-2003 cdna clone:j013092F09, full insert sequence.	
REFERENCE AUTHORS		<p>1 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Koijima,K., Namiki,T., Ohneda,E., Yabagi,W., Suzuki,K., Li,C., Ohsumi,K., Shisiki,T., Foundation of Advancement of International Science Genome Sequencing &amp; Analysis Group, Oono,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Naitaka,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Mitsu,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J., Kusumegi-T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN: Kanai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Inii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M., and Haysheiziaki,Y. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice Science 301 (5631), 376-379 (2003)</p>		
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS		japnica rice Science 301 (5631), 376-379 (2003) 2752273 12869764 2 (bases 1 to 2045) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanegaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayaehizaki,Y., Hayatsu,N., Hizmoto,K., Hirooka,T., Hoti,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,U., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Koijima,K., Koijima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara.C., Kurosaki.T., Kusumegi,T., Li,C., Lu,M., Masuda.H., Matsubara.K., Matsuyana,T., Miura.J., Miyazaki.A., Mizuno,K., Morikami,K., Murata.M., Nagata.T., Nakamura.M., Namiki.T., Narikawa.R., Nilkura,J., Nishi.K., Nomura.K., Nunaseki.N., Ohneda.E., Onho.M., Ohsuki.K., Oka.M., Ooka.H., Osato.N., Ota.Y., Osono.Y., Ryu.R., Satoh.H., Sakai.C., Sakai,K., Sakazume.N., Sano H., Sasaki.D., Sato.K., Satoh.K., Shibata.K., Shinagawa,A., Shiraki.T., Shisiki.T., Sogabe,Y., Sugano.S., Sugiyama,A., Suzuki.K., Suzuki.Y., Tagami,M., Tagami-Takeda.Y., Tagawa,A., Takahashi.F., Takaku-Akahira,S., Tanaka.J., Tomaru,A., Toya,T., Tsumoda.Y., Ueda,M., Waki,K., Xie.Q., Yabagi.W., Yamada.H., Yamamoto.M., Yaunishi.A., Yazaki,J., Yokomizo,S. and Yoshimura,A. Direct Submission		
TITLE				

JOURNAL

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28k full-length cDNA clones from japonica  
rice.

COMMENT

URL: <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J.,  
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, S., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otonari, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
Mizuno, K., Naitikawa, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S.,  
Sugiyama, A., Suzuki, Y., Tsumoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
Yoshimura, A., Matsubara, K. and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kigawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,  
Koya, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
Sasabe, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
Takahashi-Kahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
Yasunishi, A. and Hayashizaki, Y.

FEATURES

Location/Qualifiers

1..2045  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cuiivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J013092F09"

ORIGIN

Query Match 12.5%; Score 248.6; DB 8; Length 2045;  
Best Local Similarity 52.3%; Pred. No. 4.2e-40;  
Matches 730; Conservative 1; Mismatches 632; Indels 33; Gaps 7;

81 TGGCCACGACCTTGAATGGGCTTCGCAACGGCCGCTTACAGATGAAAGGCGCTCA 140  
256 TCCCGGAGGACCTTCATCTTCGCGACCGGCTCCCGGCTTATCAGTACGAAGCGCTGTA 315  
141 AAGAAAGTGGCGGCGGCGCTTCATCTGGGACAGTACCTGCCACTGGAGCCATCCGCA 200  
316 ACGAAGCGCGGACGAGGCGCGAGCATCTGGGACAGTACGCTCACTTCAGGCAAAAGTTG 375  
201 CCNACGCGCCCAACGCGGATGGCTTGGCATCACTACACCGCTACGATGAGGACTTTG 260  
376 AAGATGTAGCAACGCGGATGTAGCACTTTTACATCCCTACAGAGGAGATCTGA 435  
261 ATCTCTTGACCAAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGCTGGTCCGATCA 320  
436 ACTTCTGTGACCGACATGAAATGATGAGCGCTTCCTGTTCTCATTTGGAGAGAGATCC 495  
321 TTCCCTCGCGGCGGACGCTGATCCGCTCAACGAGAGGAAATGATTTTACAGCAAAAC 380  
496 TGCNAAATGAAACATCACTGAGGAAATCAACAAAGGGGATTTGCTTTCAACAGGCC 555  
381 TGAATTAGCGCCCTGTAGGCGGGGTATCAGCGCTTGGGTGATCTTTGTAACCTGGATC 440  
556 TGAATTAACGAGGTCATATCTAGAGGGTGAAGCATTTTGTCAACATCTTCCATTTGACA 615  
441 TGCCTTAGCGGCTTACAGATCGCTATGAGGCTGAGCTCAACGTAAGAGAGCTCCAGCTGG 500  
616 CCCACAGGCTCTGAGGACAAATACCGGAGCTTCTCA---GTGAAAACATTTGTGAAGG 672

QY 501 ACTTTAGCGGTATGCGAGGTTGTGCTTTGAAAGTTTGGGAGCCGATCCAGAACTGGA 560  
DB 673 ATTTCTGAGACTATGCGAGAGGTGTGCTTCCGTAAGTTCCGCGACCGGTGAAATGATGA 732  
QY 561 TCACATCAACGAGCCGATGATTCAGGCACTATGATATATGTCACCGGACGACGCCC 620  
DB 733 ACAGTTTCAACGAGCCGATGATTTCTGCGCGCGGCTTACGCGCTCCGACCAAGGCC 792  
QY 621 CCGGCGAGAGC-----ACATTTAAGACATCTCCACGAGGCAACCTGCACGTAGC 674  
DB 793 CCGGCGGTCTGCGCGTACGCTTCCAAAGAAATGCGCCCCGCGACTCCGCGAAGAGC 852  
QY 675 CCGGTCTGCTGGAAGAGCCGAGATCATGAGCATCCGCGCGGTGCGCTTACAGCA 734  
DB 853 CGTACGTCCTCCGCGCAACACTGCTCTGCGCCACCGCAGGCGCTCGCTTACCGCC 912  
QY 735 GGGACTTTGCGCCCTTGCAGAAAGGCGCAATGCGCATCTGCTCAAAGGCGACTACTATG 794  
DB 913 AAAAGTACAGGCGAGCAAGGAGGCGAGATGCGCATACGCGAGTGTGCACTGTTG 972  
QY 795 AGCCCTGGAGACGATGAGCTCGGAGACAGAGGCTGTGACGACGATGGAATTTTC 854  
DB 973 TGCCTT---ACAGCGAGCGCGCGCGGACAGACGCGCTCAGGCGCAGCTTGACTTCA 1029  
QY 855 ACATTTGCTGTTTGGCAATCCATCTTCTTGAAGAGAGCATTCAGAGAGATGAGAGA 914  
DB 1030 TGTACGAGATGTTCATGAGACCGATGTGTT---CGGCGACTTACCGGCGACATGAGA 1086  
QY 915 AGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCCATCTCAATGCCG 974  
DB 1087 AGTTGTCTGCTGACCGGCTGCGCCAAATTCACAGCGAGATGCGAGTTGTGAA---GG 1143  
QY 975 GAGAGACGACTTCTACGCGATGAAATTAATCAATCCAGTTGCGG-----C 1022  
DB 1144 GATCTTACGACTTCATGAGCTCTCACTACTACCAACAACTAGGTAAGAGGTTCTAC 1203  
QY 1023 GCCACCTTACGCGGTCCGTCGCGCGAGAGAGCACTATCTCGGCGCATTCATGAGACACAGG 1082  
DB 1204 GACGCCGAGCAAGCTCAAAACCGGCTTACGCTACTCAACATGAGGTAAATCAACAGCTT 1263  
QY 1083 AGAATTAAGACGCGACGCGCGGCTTGGGAGAGAGCGGCTCGGCTGCTGCTGCC 1142  
DB 1264 ACCGAAAGGCGGTCGCGATCGGTCCGCGGCTTTTCAAGATCTTTCACCTTACGCGC 1323  
QY 1143 CGGACATGTTCCGGAAGCATCTGCGCGGCTGTACGCGCTGTACGCGCAAGCCATCTACA 1202  
DB 1324 CGGCTCTGCGTGAAGCTGCTTCTGTATACCAAGGAAATGACATGACATGATCTACA 1383  
QY 1203 TCACGAGAAACGAGTCCGCTGCGTGGAGAGAGAACTAGACGCGAGAGAGCGCTCA 1262  
DB 1384 TTGCAAGAAACGAGACCGACGAGGCGAACAACAGCAGATCTTATCGCGAGGCGCTCA 1443  
QY 1263 ACGACCTCTTCCGATCCGATCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGA 1322  
DB 1444 AGGATGACAAACGAAATATCTTCACTACAGATCTCCGCT---TCAGTACAGCTGCA 1500  
QY 1323 CCAGAGACGCGCTGCTGCTCAAGGGTACTTTGCTGCGGCTGCTGCTGATTAATTTGAAT 1382  
DB 1501 TCNAGGAAAGGGGTAAGGTAAGGCTTACTTCACTGAGAGCTTCACTGAGAGATTTGAG 1560  
QY 1383 GGTCAAGATCTACGAGACCGAGTTGGGCTGCGTTCACAGCTACACACCTCAAGC 1442  
DB 1561 GGGGCGACGCTTACACCGCGCTTGGGCTCTATCTACGTCGACCGGAAACTCTCAAG 1620  
QY 1443 GCAAGCCCAAGAGTTC 1458  
DB 1621 GGTATCGCAAGAGTTC 1636

RESULT 10  
AF439322  
LOCUS AF439322 2068 bp DNA linear PLN 20-JUN-2002



DEFINITION Talaromyces emersonii beta-glucosidase 1 (bg1) gene, complete cds.  
 ACCESSION AF439322  
 VERSION AF439322.3 GI:21493049  
 SOURCE Talaromyces emersonii  
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Talaromyces.  
 REFERENCE 1 (bases 1 to 2068)  
 AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
 TITLE Molecular cloning and expression analysis of beta glucosidase from Talaromyces emersonii  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2068)  
 AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-OCT-2001) Biochemistry, National University of Ireland, Galway, Ireland  
 REFERENCE 3 (bases 1 to 2068)  
 AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Ireland  
 REFERENCE 4 (bases 1 to 2068)  
 AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JUN-2002) Biochemistry, National University of Ireland, Galway, Ireland  
 REMARK  
 COMMENT Sequence update by submitter  
 FEATURES  
 source  
 location/Qualifiers  
 1..2068  
 /organism="Talaromyces emersonii"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:68825"  
 <304..>2068  
 /gene="bg1"  
 /join(<304..383,472..663,732..1204,1258..1843,1930..>2068)  
 /gene="bg1"  
 /product="beta-glucosidase 1"  
 /join(304..383,472..663,732..1204,1258..1843,1930..2068)  
 /gene="bg1"  
 /EC\_number="3.2.1.21"  
 /note="glycosyl hydrolase"  
 /codon\_start=1  
 /product="beta-glucosidase 1"  
 /protein\_id="AAL34084.2"  
 /db\_xref="gi:21284363"  
 /translation="MTTETPTTTTSTLPDFLWGFATASQIEGAFDEGRPLSIM  
 DPESKTPGVKVENGTGVAACDSYHRTGEDIILKYKATYFSLSWRIILPLGRND  
 PINEKGLQYKRFLLDLHAAGIEPVTLPHEMLPDELKRYGGMNKEPFAVNYA  
 RVFNALGSKVHMTTFNBPWCSSVGLHNTGHAQRTSDRKSREGSTPEMIVGH  
 NLVAHGTIVDIYRREFKEQGEIGITLNGWAPWDPENADVEACDRKIETISW  
 PADPLVHGKVPDSWVKQLGDRLPKFTPEIARVGSNDPYGNHAKYNNVINKRGE  
 PEDIDGLDILMEDKNGNPISGETCEMLRPEPLGFKLLKMLADRYNNPKYVING  
 TSVKGESDPLIEVLNDEFRVQYRDYIGAMDAVOGVNKAVMASLIDNFEWSE  
 GRSRFGTVYVYVYKNGQRIKPKSALVIGELFNKXIRKE"  
 ORIGIN  
 Query Match 12.4%; Score 247.8; DB 8; Length 2068;  
 Best Local Similarity 53.2%; Pred. No. 6.2e-40;  
 Matches 738; Conservative 1; Mismatches 518; Indels 130; Gaps 5;  
 QY 118 CTACCAATCGAAGGCGCGCTCAAAAGAGTGGCGCGCGCTCCATCTGGGACACGTA 177  
 Db 472 CTACCAATCGAAGGCGCGCTCAAAAGAGTGGCGCGCGCTCCATCTGGGACACGTA 531  
 QY 178 CTGGCCACTGGAGCCTATCGGACCAACGCGCCCAACGCGCTGTGGATCATCTA 237  
 Db 532 CAGCAAGACCCCGCGAAGGTGAGACGAGACCAATGGCGATGTGGCTGGACCTCGTA 591  
 QY 238 CCACCGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTT 297

Db 592 CCACCGTACCGGTGAGGACATTGATATTTGAAGAAATACGGTGCAGATCTATCGCTT 651  
 QY 298 CTCCTTTGCTGTGTCG----- 313  
 Db 652 CTCGTGTCTGTGTTGTCCTGTGTATGTCGCTCTTTATTTTTCTGTGTCGTTTTTT 711  
 QY 314 -----CGATCATTTCCCTCGCGGAGCGTGCATCCGCT 349  
 Db 712 TGTCTGACGAACAATCCAGCGCATATTCCTCTCGGGGCGAAGACACCCCATTC 771  
 QY 350 AACGAGAGGGAATGATTTTACAGCAAACTGATTTAGCGCCCTGTAGAGGGGTATTC 409  
 Db 772 AATGAAAAGGCTCGAGTTCTACTCCAAAGTTCCTTGATGACTTACGCGCGCGCATTC 831  
 QY 410 AGCCCTTGGGTACTTTGTACACATGGGATCTGCTCCAGCGCTTCAAGATGCTATGGA 469  
 Db 832 GAGCCCTTGTGTCAGCTCTTCCACTGGGATCTGCGCAGACTGATGAAGGGATCGGG 891  
 QY 470 GCGTGCCTCAACGTGAGAGGTCCAGCTGACTTTGACGGGTATGCGAGTTTGCTTTT 529  
 Db 892 GCGATCTCAACAGAGAGGAGTTGTGGCCGACTACGCACTATGCCGTGTCGTTTC 951  
 QY 530 GAACGTTTGGGACCGGATCCAGAACTGATTCACATGAAGGCCCTGGATTCAGGCC 589  
 Db 952 AACGCGCTCGGCTCGAAGGTCAAGCACTGATCTACTTTAAGAGGCTGTGTGCACTTC 1011  
 QY 590 ATCTATGATATGCAACCGGACAGCAACGCGCGGAGAGAGCAGCTTAACAAGCACTCC 649  
 Db 1012 GTCTGAGCCACAAACCGGACAGACGCGCGCGGAGAGAGAGATGAGCAAGAAAT 1071  
 QY 650 ACCGAGGCAACACTGCCACTGAGCGGTGCTGCTGGAAGGCCCAATCATGAGCCAT 709  
 Db 1072 CCGAGGAGAACGCGACTGCGCAACCTTGATGTGTGATCAACAATTCGTGCTCAT 1131  
 QY 710 GCCCGCGCGCTGCGCTGCTACAGCAGGACTTTGCGCCCTCGCAAAAGGCCAGATCGGC 769  
 Db 1132 GGTACCGTTGTGACATCTATCGAAGAGTTCAAAAGAAACAAGCGGGGAAATTCGA 1191  
 QY 770 ATCTGCTCAAC----- 781  
 Db 1192 ATCACTCTCAACGATGAGCTCGCGCTCATCTGATTCGATTTATGTTCTAACCGCATCT 1251  
 QY 782 -----GGCGACTACTATGAGCCCTGGGACAGCAATGAGCTCGGACAAAGAGCTGCTG 836  
 Db 1252 ATCTAGCGGACTGGCGCGAACAATGGGATCCGAAACCAAGAGATGCTGAAGCTCGCG 1311  
 QY 837 AGCGAGGATGGAATTTCAATGGCTGCTGTTGCCAATCCCATCTTCTGAAGAAGACT 896  
 Db 1312 ACCGAAAGATCGAATTCGCAATCTGCTGTTCCGGAGACCGGATCTAC---CAGGGAAAT 1368  
 QY 897 ATCCAGAGAGCATGAAGAAGCAAGCTGGGCGAGAGGCTTCAAGCCCTCACTCCCGGACT 956  
 Db 1369 ACCGGAAGATCGATGTGAACAAGTTGGGCGACCGTCTGCGAAGTTACCGCCGAAAGAA 1428  
 QY 957 TTGCGATCTCTAATGCCGAGAGACGCACTTTACAGCAATGAATTAACAATCCAGT 1016  
 Db 1429 TCGCCTTGTGCAAGGAGC---AAGCATTTTATGCGCATGAACCACTATCGCAAAACT 1485  
 QY 1017 TCGCGCGCACTTACAGGCTCCGCTCCCGAGACGACTATCTCGGCGCATTCATGAGC 1076  
 Db 1486 ACATTCGCAACCGGACCGGCAACCAACCGGAGACATTTGCTGGCAACTGAGACATTC 1545  
 QY 1077 ACCAGAGATTAAGAGCGGACGCGCCGTTGGCGAGAGAGAGGCGCTCGCTGTGGCT 1136  
 Db 1546 TGATGAGAGCAAGAACCGGCAACCCCATTCGCGCCCGAGAACCAATGCGAGTGGCTGGC 1605  
 QY 1137 CTGCGCGGCAATGTTCCGGAAGCATCTCGCGGGGTATACGCTGATCGGCAAGCCC- 1195  
 Db 1606 GCTTCCGCTGCGGTTCAGAGAGCTGTAAGTGGCTTTCGAGACCGGATCAACAACCCCA 1665  
 QY 1196 --ATTAATCATCAGAGAACGATGCTCCGCTCTGAGAGAGAACTGACGTGCGAGG 1253



Db 166 AGATCTAGTACCGAGACGAGCGCTCAAGGGCGAGAGCCGCTCGAGG 1725  
Oy 1254 AGGCGCTCAACGAGCCCTTCGCGATCGGTAATTTGACTGCACTTGATGATTTCCA 1313  
Db 1726 AGGTGCTCAACGAGTTCGGGTGCACTACCGGACATATCGCGCATGTGG 1785  
Oy 1314 AGGCGATTTACCGAGAGCGGCTGCTCAAGGGGTAATTTGCGTGGCGCTTGCTGATA 1373  
Db 1786 ATGCTGTGCAAGATGGGCTCAAGCTTACATGCGTGTGAAGTTGTTAGAGT 1845  
Oy 1374 ACTTGA 1380  
Db 1846 AAGTGA 1852

RESULT 11  
LOCUS AX654334 1404 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 4204 from Patent WO03000898.  
ACCESSION AX654334  
VERSION AX654334.1 GI:29157148  
KEYWORDS  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Bristlegrass; Oryzae; Oryza.  
REFERENCE 1  
AUTHORS Chang, H. S., Chen, W., Cooper, B., Glazebrook, J., Goff, S. A., Hou, Y. M.,  
Kaeberli, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T., and Zou, G.  
TITLE Plant genes involved in defense against pathogens  
JOURNAL Patent: WO 03000898-A 4204 03-JAN-2003;  
SYNOPSIS Syngenta Participations AG (CH)  
FEATURES  
source 1..1404  
/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

ORIGIN  
Query Match 12.4%; Score 246; DB 6; Length 1404;  
Best Local Similarity 54.5%; Pred. No. 1.5e-39;  
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

Db 497 ACAGGCTGAAGCACTGATATCAGCTCAAGAGCCGACAGGTGGCCATCCAGGGCTACG 556  
Oy 603 CCACCGGCAACAGCCCGCGGAGAGACGATTA-----ACAAGACTCCACGAGG 656  
Db 557 ACCGAGGCTCCAGGCGCCCGGCGCTGCTCGTGTCTCCACCTCTACTGGAAGCGG 616  
Oy 657 GCAACATGCTCACTAGAGCGGTGCTGCTGAAAGGCCAGATCATAGAGCATGCCGCG 716  
Db 617 GCAACTTCGGACACGAGCCCTAGCTGTGCTGCCACCATTTCACTTCGCCGAGCGCGG 676  
Oy 717 CCGTGGCCGTCTACAGAGGACTTTGCGCCCTCGGAAAGGGCGAGTGGCATCTCGC 776  
Db 677 CCGCCAGCATCTACAGGACAAATACAGGCGAGCAGAAAGGCGAGCTTGGATAGCGT 736  
Oy 777 TCACGCGCATCTATGAGCCCTGGGACAGCATATAGCTCGGACCAAGAGGCTGCTG 836  
Db 737 TCAGCTGATGTGTTCGAGCCGATGTCCAAACACAGAT--CGACATCGAGGCGGCA 793  
Oy 837 AGGACGATGGAATTTCAATGAGCTGTGTTGCCAATCCATCTTCTGAAAGAGACT 896  
Db 794 AGAGAGCGAGAGGTTTCACTAGATGATGTTTCTGATCCGTTCTT---CGGCGACT 850  
Oy 897 ATCCAGAGAGCATGAAGAAAGACAGCTGGCGAGAGGCTTCCAGCTTCCGCGAGCT 956  
Db 851 ACCGCGCAGATGAGGCGAGGCTGGGGAGAGGCTGCCAGGTTCAAGCGGATGAGG 910  
Oy 957 TTGCGATCTCTCAATGCCGAGAGACCGACTTCTACGCGATGAATTTCTACATCCCACT 1016  
Db 911 CCGCGCTGTGCAAG---GGGCGCTGGATTTCTGTGGCTAAACCTACACCACTACT 967  
Oy 1017 TCGCGCGCCACCTAGACGAGTCCGCTCCCGAGA-----CGACTATCTGGGCGCCATCCA 1071  
Db 968 ACAAGAGGACAAACAAACCAACATATCGGACATTTGCTCAACAAACCTTGGCAGCA 1027  
Oy 1072 TGAGCACAGAGAGAT-----AAGACGCGAGCCCGTGGCGAGAGAGCGGCTCG 1124  
Db 1028 CCGGACCGCTCAGACCTCCCATTCAGAAATGGGAAGCAATTGGAGATAGGGCAATTGCA 1087  
Oy 1125 CCGTGTGCGCTCTCTCCCGGAGATGTTCCGGAAGATCTCGCCGCGGTATAGGCTGT 1184  
Db 1088 TATGCTGTATCATTTGCCCCGAGGATAGAGGCTGTATGAACTATGTCAAGAAAGT 1147  
Oy 1185 ---ACGCAAGCCCATCTATCATACCGAGAGAGGATGCGCGTCCCTGGAGAGAGACA 1241  
Db 1148 ACAACAGCCACCAAGTATCATCTGAATAAGGATGATGACAGCAACACCCGTTCA 1207  
Oy 1242 TGAAGTGCAGAGAGCGCTCAACGAGCCCTTCGCAATCGGTAATTTGAATCGCACTTG 1301  
Db 1208 TTTTCATCAAGAGAGCGCTCTCAAGAGACAGCAAGAGATCAAAATACCAATGATACCTCA 1267  
Oy 1302 ACTCGATTTCCAGAGCCATTTACCAAGAGAGCGCTGCTGCTCAAGGGGTACTTTGCGTGG 1361  
Db 1268 CCAATCTGTGCTTTCATCATCAGAGAGAGAGGATGATGATGATGATGATGATGATGAT 1327  
Oy 1362 CGTTGCTGATTAATTTGAA--TGGTGAATGCTTACGAGCCGATTTGCGCGTCAAGTTG 1420  
Db 1328 CTCTGCTGACATCTGGAATTTGGCGCGGATTAATCTCTGAGATTTGGGCTCTACTTC 1387  
Oy 1421 ACAGACTTACA 1430  
Db 1388 GTGAGCTTACA 1397

RESULT 12  
AK067231 2148 bp mRNA linear PLN 24-JUL-2003  
LOCUS AK067231  
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013095012, full  
insert sequence.  
ACCESSION AK067231  
VERSION AK067231.1 GI:32977249  
KEYWORDS F11 cDNA; CAP trapper.  
SOURCE Oryza sativa (japonica cultivar-group)

## ORGANISM

Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriatroidae; Oryzaceae; Oryza.

## REFERENCE

1 The Rice Full-length cDNA Consortium. National Institute of  
Agricultural Sciences Rice Full-length cDNA Project Team,  
Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K.,  
Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Oka H., Hotta I.,  
Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C.,  
Ohtsuki K., Shishiki T., Foundation of Advancement of International  
Science Genome Sequencing & Analysis Group, Ootomo Y., Murakami K.,  
Iida Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y.,  
Kurosaki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M.,  
Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Nishikura U.,  
Ikedo R., Ishibiki J., Kawamata M., Yoshimura A., Miura J.,  
Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN,  
Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S.,  
Hara A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M.,  
Kagawa I., Kondo S., Kono H., Miyazaki A., Ootomo Y., Ota Y.,  
Sato R., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T.,  
Yoshino M., and Hayashizaki Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice  
Science 301 (5631), 376-379 (2003)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
japonica rice  
Science 301 (5631), 376-379 (2003)  
2 (bases 1 to 2148)  
Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Doi K.,  
Fujimura T., Fukuda S., Hanagaki T., Hara A., Hashizume W.,  
Hayashizaki Y., Hayatsu N., Hiramoto K., Hirooka T.,  
Hori F., Hotta I., Iida J., Iida Y., Ikeda R., Imamura K.,  
Imotani K., Ishibiki J., Ishii Y., Ishikawa M., Itoh M., Kagawa I.,  
Kawagawa S., Katoh H., Kawagashira N., Kawai D., Kawamata M.,  
Kikuchi S., Kishikawa-Hirozane T., Kishimoto N., Kobayashi M.,  
Kodama T., Kojima K., Kojima Y., Kondo S., Kono H., Kouda M.,  
Koya S., Kurihara C., Kurosaki T., Kusumegi T., Li C., Lu M.,  
Masuda H., Matsubara K., Matsuyama T., Miura J., Miyazaki A.,  
Mizuno K., Murakami K., Murata M., Nagata T., Nakamura K.,  
Namiki T., Narikawa R., Nishikura U., Nishi K., Nomura K.,  
Numasaki R., Ohneda E., Ohno M., Ohtsuki K., Oka M., Ooka H.,  
Ootomo Y., Ota Y., Ootomo Y., Ryu R., Satoh H., Sakai C., Sakai K.,  
Sakazume N., Sano H., Sasaki D., Sato K., Satoh K., Shibata K.,  
Shinagawa A., Shiraki T., Shishiki T., Sogabe Y., Sugano S.,  
Sugiyama A., Suzuki K., Suzuki Y., Tagami M., Tagami Takeda Y.,  
Tagawa A., Takahashi F., Takaku-Akahira S., Tanaka T., Tomaru A.,  
Toya T., Tsunoda Y., Ueda M., Waki K., Xie Q., Yahagi W.,  
Yamada H., Yamamoto M., Yasunishi A., Yazaki J., Yokomizo S., and  
Yoshimura A.  
Direct Submission  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
Agricultural Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica  
rice.  
URL: <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi S., Satoh K.,  
Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J.,  
Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T.,  
Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T. and  
Yamamoto M.  
PAIS Genome Sequencing & Analysis Group: Ootomo Y., Iida Y.,  
Fujimura T., Ikeda R., Ishibiki J., Kawamata M., Kobayashi M.,  
Kodama T., Kurosaki T., Kusumegi T., Lu M., Masuda H., Miura J.,  
Mizuno K., Narikawa R., Nishikura U., Oka M., Ryu R., Sugano S.,  
Sugiyama A., Suzuki Y., Tsunoda Y., Ueda M., Xie Q., Yokomizo S.,  
Yoshimura A., Matsubara K. and Murakami K.  
Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi J., Aizawa K.,  
Akimura T., Arakawa T., Carninci P., Fukuda S., Hanagaki T.,  
Hara A., Hashizume W., Hayashizaki K., Hayatsu N., Hiramoto K.,

Hirooka T., Hori F., Iida J., Imamura K., Imotani K., Ishii Y.,  
Itoh M., Kagawa I., Kanagawa S., Katoh H., Kawai J.,  
Kishikawa-Hirozane T., Kojima Y., Kondo S., Kono H., Kouda M.,  
Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M.,  
Nakamura M., Nishi K., Nomura K., Numasaki R., Ohno M., Ootomo Y.,  
Ota Y., Satoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T.,  
Sogabe Y., Tagami M., Tagami Takeda Y., Tagawa A., Takahashi F.,  
Takaku-Akahira S., Tanaka T., Tanaka T., Tomaru A., Toya T., Waki K.,  
Yasunishi A. and Hayashizaki Y.  
Location/Qualifiers  
1. 2148  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/DB\_xref="taxon:39947"  
/clone="J013095012"

## ORIGIN

Query Match 11.1%; Score 221.6; DB 8; Length 2148;  
Best Local Similarity 55.9%; Pred. No. 1,3e-34;  
Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5;

81 TGCCCAACGACTTGAATGGGCTTCGCAACGGCCGCTTACGAGTGAAGGCGGCTCA 140  
|||  
203 TCCCGAGAGACTTCAATCTCGGACCGGCTCCGCGCTTATCATGACGAAGGCGCTGTGA 262  
|||  
141 AAGAGGTGGCCGCGCGCCCTCCATCTGGGACACGATCCGACCTGGAGCATTGGCGCA 200  
|||  
263 ACGAAGCGCGCAAGGGCCGAGATCTGGGACGATGCTCATCATTCACGCAAACTTG 322  
|||  
201 CCAAGCGCGCAAGCGGATGGCTTGCATCTACTACCAACGCTACGATGAGACTTTTG 260  
|||  
323 AAGATGTACCAACGGGAGATGTACGATGAGACTTTTACATCGCTACAAAGAGATGTGA 382  
|||  
261 ATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGCTCGGATCA 320  
|||  
383 ACTTCGTGACCGACATGAAATGAGACCGCTTCTCATTTGCTTGGAGCAGGATTC 442  
|||  
321 TTCCTTCGCGCGGAGGCTGATCCGCTCAACGAGGAGGAATGATTTTACAGCAAC 380  
|||  
443 TGCCAAATGAACCATCATGATGAGGAATCAACAAAGGGGATTTGCTTTTCAACAGCC 502  
|||  
381 TGATTGACGCTTGTAGGCGGGGTATCAAGCGCTTGTGGTGAATCTTTTACCATGGATC 440  
|||  
503 TGATTACAGAGTATATCTAGAGGTTGAAGCATTTGTACCATCTTCATTTGACA 562  
|||  
441 TGCTCAGCGCTTACGATGCTATGAGGCTGGCTCAACGTAAGAGAGTCCAGCTGG 500  
|||  
563 CCCACAGGCTCTGAGAGACAAATACCGAGCTTCTCA---GTGAAACATGTGTGAAG 619  
|||  
501 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTGA 560  
|||  
620 ATTTCGTGACTATGCGAGCGTGTGCTTCTGCTGTGAGTTCCGCGACCGGGTGAAGTATGA 679  
|||  
561 TCACCATCAACGAMCCCTGTGATTCAGGCATCTATGATATGCCAGCGGACGAACGCC 620  
|||  
680 ACAAGTTCAAGACCGCATGATCTTGTGCGCGGCGGCTAGCGCTCCGACCAAGGCC 739  
|||  
621 CGGCGAGAGC-----AGATTAAACAGACTCCACGAGGAGAAACATCGCCACTGAGC 674  
|||  
740 CGGCGCGTGTCTCGCGTATCTTCAAGAAATCGCCCCCGGAGACTCGGCAACGAGC 799  
|||  
675 CGTGGCTGCTGGAAGGCCAGATCTAGAGCCATGAGCCGCGCGCGCTGAGTCAAGCA 734  
|||  
800 CGTAGCTGCGCGGCAACAATCTGCTGCTGCGCCACGCGGAGCCGCTCCGCTTACCGCC 859  
|||  
735 GGGAGCTTTCGCGCTCGAAAGGCGCAGATGCGCATCTGCTCAACGCGCATCTATG 794  
|||  
860 AAAAGTACCAAGCAGGAGAGAGGCGAGATCGGCATCAACGAGGTGTGCGACTGTTCG 919  
|||  
795 AGCCCTGGACAGCAATGAGCTCTGGGACAAAGAGGCTGCTGACGACGAGTGAATTC 854  
|||

Db 920 TGCCCT--ACAGCAGCGCCGCGCAGACGCCCTCAGGCGCGCCTCGACTTCA 976

Qy 855 ACATGGCTGTGTTGGCATTCCTCTTGAAGAAGCACTATCCAGAGAGATGAGA 914

Db 977 TGACCGATGTCATGAGACCGCATGTGTT---CGCGCATACCGGGGCACTAGGGA 1033

Qy 915 AGCAGCTGGCGGAGAGCTTCACGCCCTCACTCCCGGAGCTTGGCATTCATGCCG 974

Db 1034 AGTTGGTCGGTGACCGGCTGCCCAATTCACAGCCGAGCGAGTGGTGTGA---GG 1090

Qy 975 GAGAGCCGACTTCTACGGCATTAATTACTACATCTCCAGTTGCC 1020

Db 1091 GATCCTACGACTTCATCGCGCTCACTACTACACCACTACGCG 1136

RESULT 13

AK120430 2148 bp mRNA linear PLN 29-OCT-2003

LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013095M12, full

DEFINITION Insert sequence.

ACCESSION AK120430

VERSION AK120430.1 GI:337990053

KEYWORDS FLI\_CDNA; CAP trapper.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Eriarthroideae; Oryzaceae; Oryza.

REFERENCE 1

AUTHORS The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team;

Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohta,Y., Onishi,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ootomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nikiura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Mura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matubara,K., RIKEN; Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashidume,M., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Oota,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M., and Hayashizaki,Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 22752273

PUBMED 12869764

REFERENCE

AUTHORS

TITLE Collection, mapping, and annotation of 28k full-length cDNA clones from japonica rice

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 2148)

AUTHORS Kikuchi,S.

TITLE Direct Submission

SOURCE Submitted (31-JAN-2003) Shoeni Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 32k full-length cDNA clones from japonica rice.

URL: http://cdna01.dna.affrc.go.jp/cDNA/

NIAS Rice full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohta,Y., Onishi,K., Shishiki,T., Yamamoto,M., and Nakahama,Y.

FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Mura,J., Mizuno,K., Narikawa,R., Nikiura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K., and Murakami,K.

Genome Exploration Research Group in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohtomo,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A., and Hayashizaki,Y.

Location/Qualifiers

1. 2148

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultiyar="Nipponbare"

/db\_xref="taxon:39947"

/clone="J013095M12"

ORIGIN

Query Match 11.1%; Score 221.6; DB 8; Length 2148;

Best Local Similarity 55.9%; Pred. No. 1.3e-34;

Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5;

Qy 81 TGCCCAACGACTTGAATGGGGCTTCGCAACGCGCCCTACAGATCGAAGCGCGCTCA 140

Db 203 TCCCGAGAGACTTCATCTTCGGACCGCGCTCGCGCTTATAGTAGAAGCGCTGTGA 262

Qy 141 AAGAAGGTGCGCGCGCGCTTCATCTTCGAGACGTAATGCACTGGAACCATCGCGCA 200

Db 263 ACGAAGCGCGCGAGAGCGCGAGCATCTGGAGACAGTACCTCCATTCAGGCAAGTTG 322

Qy 201 CCAAGCGCGCGCAACCGCGAGTGGCTTGGATACAGACCGCTCGATGAGAGACTTGG 260

Db 323 AAGATGTGACGACCGGGAGTAGACGTGACATTTTACACATCGCTACAGAGAGATCTGA 382

Qy 261 ATCTCTTGACCAAGTACGCGCGCAAGAGCTTACCGCTTCTCTTGTGTGTCGCGATCA 320

Db 383 ACTTCGTGACCGACATGACATGAGAGCGCTTCGCTTTCATGCTGTGAGAGAGATCC 442

Qy 321 TTCCCTCGCGCGCGCGCTGATCCGTCACAGAGAGGAATTGATTTTACAGCAAAC 380

Db 443 TGCCAAATGAAACCATCACTGAGAGGAATCAACAAGAGGGGATTTGCTTTTACACAGCC 502

```

QY 381 TGAATGACCCCTGTTGAGGCGGGGTATCGCCCTTGGTGAATCTTGTACACAGGGATC 440
DB 503 TGAATTAACGAGGTCATATCTAGAGGGTTGAAGCCATTGTGCACATCTTCCATTTTGACA 562
QY 441 TGGCTGAGGCGCTTCCAGATCGCATGAGAGGCTGCTCAACGTGAAGAGAGGTGAGCTGG 500
DB 563 CCCACACAGGCTCTGAGAGACAATAACGGAGCTTCCCA---GTGAATAATGTGAAGG 619
QY 501 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCCAGTCCAGAACCTGGA 560
DB 620 ATTTTCGAGCTATGCGAGCGTGTGCTTCCGTAGTTCCGCGAGCCGGGTGAAGTCAAGA 679
QY 561 TCCCATCAACGAGCCCTGATTCAGGCACTTATGATATGACACCGGAGCAACGCCCC 620
DB 680 ACGGTTCAAGAGCCCATATCTTCTTCCGCGCGGCTACGGCTCCGGGACCAAGGCC 739
QY 621 CGGCGAGAGAGC-----AGCATTAACAAGCACTTCACCGAGGGCAACACTGCACTGAGC 674
DB 740 CCGGCGCTGCTGCGCGGTAGCGTCCAAAGAAATCGCCCCCGGCGCACTCCGGCAACGAGC 759
QY 675 CGTGGCTGCTGGAAGAGCCCAATCATAGAGCCATGCCCCGCGCTGGCCGTCTTACAGCA 734
DB 800 CGTACGTCGCGCGGCAACCTGCTGCTGCCCAAGGCGCGTCCGCTTACCGCC 859
QY 735 GGGAGCTTTCGCCCTCGCAAAAGGGCCAGATCGGATCTGCTCAACGCGCATCTATG 794
DB 860 AAAAGTACGAGGCAACGAGAGGAGATCGGATCAACGAGTGTGCGACTGCTTG 919
QY 795 AGCCCTTGGGACAGCAATGAGCCTCGGAGCAAGAGGCTGCTGAGCGACGATGGAATTC 854
DB 920 TGCCCT---ACAGGAGCGCGCGCCGCCCAAGCAAGCGCGTCAAGCCTTCACCTTCA 976
QY 855 ACATTTGCTGTTGGTCCCATTCCTCTTGAAGAAGAGATATCCAGAGCATGAGA 914
DB 977 TGTAGGATGTTATGAGCCCGATCGTGT---CGGAGATACCGGCGCATTAAGA 1033
QY 915 AGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTCGATCTCTCAATGCG 974
DB 1034 AGTTGTGTGATGACCGGCTGCCAAATTCACAGCGAGCACTCGAGATGTGTGA---GG 1090
QY 975 GAGAGACCGACTTCTACGGCATGAAATCTACATCCCAATGCTGGC 1020
DB 1091 GATCTCACTGACTTCACTGCGCTCAACTACTACACCACTACGCG 1136

RESULT 14
AK066710
LOCUS
DEFINITION
ORyza sativa (japonica cultivar-group) cDNA clone:J013074A20, full
insert sequence.
ACCESSION
AK066710
VERSION
AK066710.1 GI:32976728
KEYWORDS
FLI_CDNA; CAP trapper.
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzaeae; Oryza.
REFERENCE
1
The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,M., Suzuki,K., Li,C.,
Ohnuki,K., Shisiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group; Otsu,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nishikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN;
Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arai,K., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,

```

Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Otsu,N., Ota,Y.,  
 Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,  
 Yoshino,M. and Hayashizaki,Y.  
 collection, mapping, and annotation of over 28,000 cDNA clones from  
 japonica rice  
 Science 301 (5631), 376-379 (2003)

JOURNAL  
 MEDLINE  
 PUBMED  
 12869764  
 2 (bases 1 to 1811)

REFERENCES  
 Adachi,J., Aizawa,K., Akimura,T., Arai,K., Carninci,P., Doi,K.,  
 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashidume,W.,  
 Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hitamoto,K., Hiraoka,T.,  
 Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,  
 Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,  
 Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,  
 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,  
 Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
 Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,  
 Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,  
 Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,  
 Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,  
 Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,  
 Otsu,N., Ota,Y., Otsu,Y., Otsu,R., Saitoh,H., Sakai,C., Sakai,K.,  
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Sugano,S.,  
 Shingawa,A., Shiraki,T., Shisiki,T., Sogabe,Y., Sugano,S.,  
 Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,  
 Tagawa,A., Takahashi,F., Takaku-Akashira,S., Tanaka,T., Tomaru,A.,  
 Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,  
 Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and  
 Yoshimura,A.

TITLE  
 JOURNAL  
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
 Agrobiological Sciences, Department of Molecular Genetics, Head of  
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
 Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT  
 This clone is one of the 28k full-length cDNA clones from japonica  
 rice.

URL: http://cdna01.dna.affrc.go.jp/cDNA/  
 NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,  
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,  
 Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,  
 Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohtsuki,K., Shisiki,T. and  
 Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y.,  
 Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,  
 Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,  
 Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S.,  
 Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,  
 Yoshimura,A., Matsubara,K. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center  
 and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,  
 Akimura,T., Arai,K., Carninci,P., Fukuda,S., Hanagaki,T.,  
 Hara,A., Hashidume,W., Hayashida,K., Hayatsu,N., Hitamoto,K.,  
 Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,  
 Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,  
 Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Otsu,N.,  
 Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,  
 Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,  
 Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,  
 Takaku-Akashira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,  
 Yasunishi,A. and Hayashizaki,Y.

Location/Qualifiers  
 1..1811  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultiar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="J013074A20"

FEATURES  
 source

ORIGIN

Query Match 10.8%; Score 215.8; DB 8; Length 1811;  
 Best Local Similarity 51.3%; Pred. No. 1.9e-33;  
 Matches 714; Conservative 1; Mismatches 648; Indels 30; Gaps 8;

QY 81 TGCCTCAACGATTGGAATGGGCGCTTCGCAACGGCCGCTTACAGATGGAAGGCGCGTCA 140  
 DB 135 TCCCAAGGAGCTTCATCTTCGGGACAGAGTTACAGAGCTTATCAGTATAGAGGCGCTTACA 194  
 QY 141 AAGAAGGTGCGCGGCGCGCTTCATCTGGGACAGTACGTGCGCATCTGAGAGCATTCGGCA 200  
 DB 195 AAGAAGGGGCAAGAGTCTTAGCGCTTGGGACAACTTCACTCACTTACAGTAAATTTT 254  
 QY 201 CCAAGCGCGCAACGGCGATGTGGCTTGCATCACTACCAACCGCTACGATAGAGACTTTG 260  
 DB 255 TAAACAATGATTAACGGGATGTGGCAATGACTTCTATCAACCATACAAAGAGATGTGA 314  
 QY 261 ATCTTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTGTGTGTGGTGGATCA 320  
 DB 315 GCCTCTGGAAGACATGAACATGATGCTTCCGGTCTCCATTTGCGTGAACAGATCC 374  
 QY 321 TTCCCTCGCGGCGGAGCTGATCCGTCACAGAGAGGAATGAGTTTACAGCAAC 380  
 DB 375 TGCCAAATGATCTTATAGTGAAGAAATAAACAAAGAGGCTTCTTCTCAACACGT 434  
 QY 381 TGAATGACGCTTGTGAGCGGGGTATCAAGCTTGGGTGACTTTGTACACTGGATC 440  
 DB 435 TGATCATGATGTCATGCAAAAGGATGATCCATTTGTCACTATCTTCACTGGGACA 494  
 QY 441 TGCCTCAGCGCTTACAGATCGCTATGAGAGCTGGCTCAACGTGAAGAGTCCAGCTGG 500  
 DB 495 CCCCCTTGGCTTGAAGAAAGAAATACGAGAGATTCCTCA--GTGAAGACATAGTGAAG 551  
 QY 501 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAACCTTTTGGGACCGAGTCCAAACTGGA 560  
 DB 552 AATACTGGAAGTTCGGAGGAGTGTCTTTCGCAAGTTCGGAGCGGTGTCAAGATCTGGA 611  
 QY 561 TCACCATCAACGAMCCCTGATTCAGGCCATCTATGATATGACACGGGACGCAACGCC 620  
 DB 612 CCACATTTAATGAGCCATTCAATATAGCCCTTACGGCTACGCAAGGCGCTGTTCCG 671  
 QY 621 CCGGAGGAGACGCA-----TTAACAAGCACTCCACGAGGCGCAACTGCGCATGAGC 674  
 DB 672 CCGGAGCATGCTCTTCAATATGTTTCCAAATCATGCGCGCTGCGGATCTCAAGTGCAGC 731  
 QY 675 CGTGGCTCGGTGAAGAGCCCAATCATAGAGCCATGCCCCGCGTGGCGCTTACAGCA 734  
 DB 732 CTAACCTCGGGACACACATCATCACTCTCCACGCTGCGCTGTCCAGCTTTACCGCA 791  
 QY 735 GGAAGCTTTCGCCCCCTCGCAAAAGGGCCAGATCGGCATCTGCTCAAGCGGCACTACTATG 794  
 DB 792 CCAAGTACAGCCCAACAGAGGAGAGATCGGCATGTGTGTGTATACCACTGTGTTG 851  
 QY 795 AGCCTTGGGACAGCAATGAGCTTCGGACAAGAGGCTGTGACGACGAGATGGAATTTTC 854  
 DB 852 TGCCTGACGACCACTCCGAGCT--GACCGTGGCGCTGTGCAACGAGGCTTACATTTCA 908  
 QY 855 ACATGTGCTGTGGTCCCAATCCCATCTTCTTGAAGAGACTATCCAGAGCATGGAAGA 914  
 DB 909 TCTATGGGTGTTCATGAGACCTTA--TCGTGCAATGTGTACTACCCAGGACCATGTGAG 965  
 QY 915 AGCAGCTGGGAGAGAGCTTCAGACCTCACTCCCGGGAAGTTGGCATCTCTCAATCCG 974  
 DB 966 GTTGGCTCGGCATTCGGCTGCGAGATTCAAGCTTGAACAGTGGCGCATGTGTGAAGGCT 1025  
 QY 975 GAGAGACGACTTCTACGGCATGATTAATACATCCAGTTTCGCGGCACTTAGACG 1034  
 DB 1026 CTTA---GCACTTATCGAGATTAACTATACACCACTACTAGAGTATACCGC 1082  
 QY 1035 GTCCCGTCCCGAGAGG-----ACTATCTCGCGGCACTTCATGAGACCAAGAGATA 1088  
 DB 1083 CGCTTAATCTCAAGAGACTATCTTACGACCTTGAACCGGCGCAACACACCGGCTTCC 1142  
 QY 1089 AGGAGCGGAGCCCGCTTGGGAGAGAGAGGCGCTCGCTGCGCTCTGCGCGGACA 1148

DB 1143 GTAATGGCAAAACCCATCGGTCCACAGGAATTTACACCGATCTTCTTCACTACCTCCAG 1202  
 QY 1149 TGTTCGGGAAGCATTCGCCCGGGGTGTAGCGCTGTAGCGCAAGCCCATATATACCG 1208  
 DB 1203 GTCTCCGTGAGCTTCCTCTTACCAAGAGAGATTAACAACCCGA---CATCTATG 1259  
 QY 1209 AGAAGGATCCCGTCCGCTGAGAGAGACATGACGTGCGAGAGGCGGTCAAGACC 1268  
 DB 1260 TTACAGAAAACGCGATCTGATGAGGTATACAGACACACTGCCAGAGCGGCTCAAGATG 1319  
 QY 1269 CTTCCGATCCGTAATCTTTGACTGCACTTGGACTGATTTTCAAGGCCATTACCGAG 1328  
 DB 1320 GACACAGAGATCGAGTTCCACTCAAGCACTCGAGTTTCGCAACAGCCAT---CAAGA 1376  
 QY 1329 ACGGCTGCTGTCAAGAGGATCTTTCGTGGCGCTTCTGATTAATCTGGAATGTGAG 1388  
 DB 1377 ATGGGCTGAACGTAAGGGCTTACTTCAAGCATTCATGAGACTGCTTGAAGTGGGTTG 1436  
 QY 1389 ATGCTACGGAACCCAGATTGCGGCTGACGTTACAGACTACACACCTCAAGCGACGC 1448  
 DB 1437 ACGGCTACCTTGACAGAGTTCCGCTCATCTACGTCAACGCAAGAGCTTACGCTTACC 1496  
 QY 1449 CCAAGAGTCTGC 1461  
 DB 1497 GCAAGAGTCTGAC 1509

RESULT 15  
 AX652977 1503 bp DNA linear PAT 22-MAR-2003  
 LOCUS AX652977  
 DEFINITION Sequence 2847 from Patent WO03000898.  
 ACCESSION AX652977  
 VERSION AX652977.1 GI:29155791  
 KEYWORDS  
 SOURCE  
 ORGANISM  
*Oryza sativa*  
*Oryza sativa*  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; *Oryza*.

REFERENCE  
 1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,  
 Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.  
 Plant genes involved in defense against pathogens  
 JOURNAL Patent: WO 03000898-A 2847 03-JUN-2003;  
 SynGene  
 FEATURES  
 Location/Qualifiers  
 1..1503  
 /organism="Oryza sativa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:4530"

ORIGIN  
 Query Match 10.8%; Score 214.6; DB 6; Length 1503;  
 Best Local Similarity 51.0%; Pred. No. 3.4e-33;  
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;

QY 81 TGCCTCAACGATTGGAATGGGCGCTTCGCAACGGCCGCTTACAGATGGAAGGCGCGTCA 140  
 DB 95 TCCCAAGGAGCTTCATCTTCGGGACAGAGTTACAGACTTATCAGTATGAGGCGCTTACA 154  
 QY 141 AAGAAGGTGCGCGGCGCGCTTCATCTGGGACAGTACGTGCGCATCTGAGAGCATTCGGCA 200  
 DB 155 AAGAAGGGGCAAGAGTCTTAGCGCTTGGGACAACTTCACTCACTTCCAGTAAATTTT 214  
 QY 201 CCAAGCGCGCAACGGCGATGTGGCTTGCATCACTACCAACCGCTACGATAGAGACTTTG 260  
 DB 215 TAAACAATGATTAACGGGATGTGGCAATGACTTCTATACCGATACAAAGAGATGTGA 274  
 QY 261 ATCTTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTGTGTGTGGTGGATCA 320  
 DB 275 GCCTCTGGAAGACATGAACATGATGATGCTTCCGGTTCCTCAATGTGTGACCAAGATCC 334

QY 321 TTCCCTCGCGGCGAGCTGGATCCCGTCAACGAGGAGGAAATTGATTTTACAGCAAC 380  
Db 335 TGCCTAAATGATTCCTTAGTGAGAGAAATAACAAAGAGGGGTGCTTTCTACAAACGCT 394  
QY 381 TGATTGACGCCCTGTGAGCGGGGTATCAAGCCTTGGGTGACTTTGTACACTGGGATC 440  
Db 395 TGATCAATGATGTATGACAAAGGATGATCCATTGTGCATATCTTCCACTGGGACA 454  
QY 441 TGCCCTCAGGCGCTTACGATCGTATGAGAGGCTGGCTCAACGTGAAAGAGGTCCAGCTGG 500  
Db 455 CCCCCTTGCTCTGAAAAAGCAAAATACGAGGATTCCTCA--GTGAAGACATAGTGAAG 511  
QY 501 ACTTTGACGGGTATGCGAGGTGTGCTTTGAAACGTTTGGGGACCGAGTCCAGAACTGGA 560  
Db 512 AATAGTGAGACTTGGGAGGTGTGCTTTCGCGAGTTCGCGACCGTGTCAAGTACTGGA 571  
QY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACGCGCAGCAACGCC 620  
Db 572 CCACCTTAAATGAGCATTCATATAGCGCTACGGCTACGGCAAGGGCGTGTGGCG 631  
QY 621 CGGCGAGAGCAGCA-----TTAACAAGCACTCCAGAGGCAACCTGCCACTGAGC 674  
Db 632 CGGGAAGATGCTCTTCAATATGTTTCCAGTCAATGCGGCTGTGACTTCAGTCGCGAGC 691  
QY 675 CGTGGCTGCTGGAAGGCCAGATCATGAGCATGCCGCGCGTGGCCGTCTACAGCA 734  
Db 692 CTTACTCTGTGGCACAACACATCCACTCTCCACGCTGCCGCTGTCCAGCTTTACGCA 751  
QY 735 GGGACTTTGCCCCCTCGCAAAAAGGSCAGATCGGCATCTCGTCAACGGCGACTACTATG 794  
Db 752 CCAAGTACAGCAACACAGAAAGGAGCAGATCGCATGTGTGTCAACCACTGTGTCG 811  
QY 795 AGCCCTGGGACAGCATGAGCCTCGGACAAGAGGCTGTGAGCGACGAGATGAATTC 854  
Db 812 TGCCCTGACACAATCCGACGCT--GACCGTGGCGCTGTGCAACGAGCCTAGACTTCA 868  
QY 855 ACATTGCTGTGTTGCCAATCCCATCTTCTTGAAAGAGACTATCCAGAGCATGAAGA 914  
Db 869 TCTATGGGTGTTCATGACCTCA--TGTCATGTGTGATCCAGGACCATGAGAG 925  
QY 915 AGCAGCTGGGCGAAGAGCTTCCAGCCCTCACTCCCGGAGCTTTGCCATCTCAATGCCG 974  
Db 926 GTTGGCTCGGCAATCGGCTGCGCAGGTTCAAGCTGAACAGTCGGCGATGTGAAGGCT 985  
QY 975 GAGAGCCGACTTCTAAGCATGAAATTACTACACATCCAGTTCCGCGCCACTAGACG 1034  
Db 986 CCTA--CGACTTTCATCGAGTTAATTACTACACACTACTACGTAAGATATACGCG 1042  
QY 1035 GTCCCGTCCCGA-----GACGGAATATCTGGCGCCATCCATGAGACACAGAGATA 1088  
Db 1043 CGCTTAATCTCAACGAGCTATCTTACGACTCGACCAACCGCGCAACACACCGGCTTC 1102  
QY 1089 AGGACGGCAGCCCGTGGCGAGAGAGGGGCTCGGCTGGCGCTCTGCGCGGACA 1148  
Db 1103 GTAAATGCAAAACCATCGGTCCACAGAAATTACACGATCTTCTTCACTACCTCCAG 1162  
QY 1149 TGTTCGGAAGCATCTGCCCCGGGTGTAGGCTGTACGCAAGCCCATCTACATCACCG 1208  
Db 1163 GTCTCCGTGAGCTCTCTCTTACACCAAGAGAGATTAACAACACCGA---CCATCTATG 1219  
QY 1209 AGAAGGATGCGCGTGCCTGAGAGAGAAATGACGTGCGAGAGGCGGCTCAAGACC 1268  
Db 1220 TTACAGAAAAAGCATGATGAGGGTAAACACAGCACATGCGCAGAGGCGCTCAAGATG 1279  
QY 1269 CCTTCGCAATCCGCTTCTTGAATCGCATTTGAGCTGCAATTTCCAGGCCATTACCAG 1328  
Db 1280 GACACAGGATCGAGTTCACTCAAAGCACTGCGAGTTGTCAACACAGGCT---CAAGA 1336  
QY 1329 ACCGCGTCTGTCAAGGGGTACTTTGCGGGCGTTGCTCATAACTTGAATGTCTAG 1388  
Db 1337 ATGGGGTGAAGTGAAGGGCTACTTCACTGGAATTCATGACTGTGTTGAGTGGGTTG 1396  
QY 1389 ATGGCTAAGGACCAATTCGGGCTCACTTCAAGACTACACCACTCCCTCAAGCGCAGC 1448

Db 1397 ACGGCTACTGTGACAGAGTTGGGCTCATCTACGTGACCGCAAGCGCTCAAGCGCTACC 1456  
QY 1449 CCAAGAACTTGCCTCGTCTCAAGACAT 1479  
Db 1457 GCAAGGATTCAGCTACTGGATCGAAGACTT 1487

Search completed: March 28, 2004, 20:51:58  
Job time : 7890.13 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 19:13:22 ; Search time 788.658 Seconds  
(without alignments)  
10724.755 Million cell updates/sec

Title: US-10-026-140-1  
Perfect score: 1991  
Sequence: 1 agccagctgcgaaccagcagc.....tttcaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N\_Geneseq\_29Jan04:\*

1: geneseqn1980s:\*\n2: geneseqn1990s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001s:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002s:\*\n7: geneseqn2003as:\*\n8: geneseqn2003bs:\*\n9: geneseqn2003cs:\*\n10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1976.6	99.3	1990	9	ADD22902 Full leng
2	1454.2	73.0	1455	9	ADD22904 T. reesei
3	598	30.0	735	3	AAFI4935 Trichoder
4	266.6	13.4	1933	3	AAFI4882 Trichoder
5	246	12.4	1404	7	ADA70881 Rice gene
6	214.6	10.8	1503	7	ADA69524 Rice gene
7	205.4	10.3	575	7	ABZ53130 Aspergill
8	182.8	9.2	1734	6	ABZ13477 Arabidops
9	182.8	9.2	1734	7	ADA67833 Arabidops
10	178.6	9.0	1533	7	ADA70381 Rice gene
11	178.2	9.0	2016	4	AA521370 Human CDN
12	178.2	9.0	2016	6	ABL88245 Human PRO
13	178.2	9.0	2016	6	ABL95734 Human ang
14	178.2	9.0	2016	7	ACD23979 Human nov
15	178.2	9.0	2016	7	ACA67120 CDNA enco
16	178.2	9.0	2016	7	ACA03729 CDNA enco
17	178.2	9.0	2016	7	ABX89267 DNA encod
18	178.2	9.0	2016	7	ACD41921 Human sec
19	178.2	9.0	2016	7	ACA04150 Human CDN
20	178.2	9.0	2016	8	ADA45772 Human hum
21	178.2	9.0	2016	8	ADA76203 Human PRO
22	178.2	9.0	2016	8	ADA18853 Human PRO
23	178.2	9.0	2016	8	ADA61476 Homo sapi

## ALIGNMENTS

24	178.2	9.0	2016	8	ADB19261	Adb19261 Novel hum
25	178.2	9.0	2016	8	ADB27802	Adb27802 CDNA enco
26	178.2	9.0	2016	8	ADA6281	Ada6281 Novel hum
27	178.2	9.0	2016	8	ADB15845	Adb15845 Human PRO
28	178.2	9.0	2016	8	ADA7631	Ada7631 Human PRO
29	178.2	9.0	2016	8	ADA67426	Ada67426 Human PRO
30	178.2	9.0	2016	8	ADB30433	Adb30433 CDNA enco
31	178.2	9.0	2016	8	ADA85729	Ada85729 Novel hum
32	178.2	9.0	2016	8	ADA69941	Ada69941 Human PRO
33	178.2	9.0	2016	8	ADA79245	Ada79245 Human PRO
34	178.2	9.0	2016	8	ADA87384	Ada87384 Novel hum
35	178.2	9.0	2016	8	ADB16586	Adb16586 Human PRO
36	178.2	9.0	2016	8	ADA91678	Ada91678 Novel hum
37	178.2	9.0	2016	8	ADB14741	Adb14741 Human PRO
38	178.2	9.0	2016	8	ADB18702	Adb18702 Novel hum
39	178.2	9.0	2016	8	ADA93917	Ada93917 Human PRO
40	178.2	9.0	2016	8	ADB19813	Adb19813 Novel hum
41	178.2	9.0	2016	8	ADB13125	Adb13125 Human PRO
42	178.2	9.0	2016	8	ACD98550	AcD98550 Novel hum
43	178.2	9.0	2016	8	ADA74379	Ada74379 Human PRO
44	178.2	9.0	2016	8	ADB24612	Adb24612 Human PRO
45	178.2	9.0	2016	8	ADA82136	Ada82136 Human PRO

RESULT 1	ADD22902	ADD22902 standard; CDNA; 1990 BP.
AC	XX	
AC	XX	ADD22902;
DT	XX	15-JAN-2004 (first entry)
DE	XX	Full length T. reesei bg15 beta-glucosidase CDNA.
XX	XX	BGL5; beta-glucosidase; fungus; ethanol; sugar;
KW	KW	fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;
KW	KW	aroma; detergent; softening agent; cotton; wood pulp; gene; ss.
OS	XX	Hypocrea jecorina.
XX	XX	
PH	XX	Location/Qualifiers
FT	XX	Key
FT	XX	CD5
FT	XX	62..1516
FT	XX	/*tag= a
FT	XX	/product= "BGL5 beta galactosidase"
FT	XX	/transl_except= (pos:98..100,aa:Gln)
FT	XX	/transl_except= (pos:551..553,aa:Gln)
FT	XX	/transl_except= (pos:572..574,aa:Xaa)
FT	XX	/note= This codon has an apparent 1 nucleotide deletion which alters the reading frame. Xaa= unknown.
US	XX	US2003114330-A1.
PD	XX	19-JUN-2003.
XX	XX	
PF	XX	18-DEC-2001; 2001US-00026140.
XX	XX	
PR	XX	18-DEC-2001; 2001US-00026140.
XX	XX	
PA	XX	(DUNN/) DUNN-COLEMAN N.
PA	XX	(GOED/) GOEDEGEBOUR F.
PA	XX	(WARD/) WARD M.
PA	XX	(YAOJ/) YAO J.
XX	XX	
PI	XX	Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;
XX	XX	
DR	XX	WPI; 2003-810909/76.
DR	XX	P-PSDB; ADD22903.
XX	XX	
PT	XX	New substantially purified BGL5 polypeptide with the biological activity





QY 1681 GCCTCTATACAGCTTATAGGCTGTGTTACTTGCCTTTTGTCTCTTCTGT 1740  
DB 1680 GCCCTTATACAGCTTATAGGCTGTGTTACTTGCCTTTTGTCTCTTCTGT 1739  
QY 1741 CTGTCTGTGCTGCTGTCTATGTGTACCTATCTGCGCTTGGCAAGATACT 1800  
DB 1740 CTGTCTGTGCTGCTGTCTATGTGTACCTATCTGCGCTTGGCAAGATACT 1799  
QY 1801 AGCAGAGTTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1860  
DB 1800 AGCAGAGTTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1859  
QY 1861 AAGGCCCATGTTAGGCTCATGTTCAATGTTATCTACATCAGCATTCACTA 1920  
DB 1860 AAGGCCCATGTTAGGCTCATGTTCAATGTTATCTACATCAGCATTCACTA 1919  
QY 1921 CCAAGCGCAGAGGCGCAGCAGAGTCTCATCTTACCTGTATATACGCTTTTAAAA 1980  
DB 1920 CCAAGCGCAGAGGCGCAGCAGAGTCTCATCTTACCTGTATATACGCTTTTAAAA 1979  
QY 1981 AAAAAAAAAA 1991  
DB 1980 AAAAAAAAAA 1990

## RESULT 2

ADD22904  
ID ADD22904 standard; cDNA: 1455 BP.

XX AC ADD22904;

DT 15-JUN-2004 (first entry)

DE T. reesei bg15 beta-glucosidase cDNA.

XX BG15; beta-glucosidase; fungus; ethanol; sugar;

KM fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;  
KM aroma; detergent; softening agent; cotton; wood pulp; gene; ss.

XX OS Hypocrea jecorina.

XX Key Location/Qualifiers

FT CDS 1..1455

FT /\*tag= a

FT /product= "BG15 beta galactosidase"

FT /tranyl\_except= (pos:37..39,aa:Gln)

FT /tranyl\_except= (pos:511..513,aa:Xaa)

FT /tranyl\_except= (pos:1219..1221,aa:Pro)

FT /note= "Xaa= unknown"

XX US2003114330-A1.

XX 19-JUN-2003.

XX 18-DEC-2001; 2001US-00026140.

XX 18-DEC-2001; 2001US-00026140.

XX (DUNN/) DUNN-COLEMAN N.

XX (GOED/) GOEDEGEBOUR F.

XX (WARD/) WARD M.

XX (YAO/) YAO J.

XX Dunn-Coleman N, Goedegebuur F, Ward M, Yao J;

XX WPI: 2003-810909/76.

XX P-PSDB; ADD22903.

XX The invention discloses a substantially purified BG15 polypeptide with  
CC the biological activity of a beta-glucosidase. Also claimed is an  
CC isolated polynucleotide derived from a fungal source, which comprises a  
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,  
CC a Trichoderma reesei bg15 nucleic acid coding sequence, or its  
CC complement, an expression construct including the T. reesei bg15 cDNA  
CC sequence, a purified recombinant enzyme having beta-glucosidase activity,  
CC a recombinant host cell comprising a deletion or insertion or other  
CC alteration in the bg15 gene which inactivates the gene and prevents BG15  
CC polypeptide production, an antisense oligonucleotide, a detergent  
CC composition comprising the polypeptide and expressing a heterologous  
CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BG15  
CC is useful for producing ethanol, which involves contacting a biomass  
CC composition with an enzymatic composition comprising the polypeptide to  
CC yield a sugar solution, adding to the sugar solution a fermentative  
CC microorganism and culturing the fermentative microorganism under  
CC conditions sufficient to produce ethanol, where the biomass composition  
CC may be optionally pretreated. The method further involves the addition of  
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with  
CC a dilute acid. BG15 is useful in wine making for enhancing the potential  
CC aroma of the finished wine product. The detergent is useful as softening  
CC agent and also useful for improving the feel of cotton fabrics and for  
CC degrading wood pulp into sugars. The sequence presented is the T. reesei  
CC bg15 cDNA.

XX Sequence 1455 BP; 312 A; 444 C; 423 G; 274 T; 0 U; 2 Other;

XX Query Match 73.0%; Score 1454.2; DB 9; Length 1455;

XX Best Local Similarity 100.0%; Pred. No. 0;

XX Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTCGTAGCTGCGCCCAAGACTTGAATGGGCTTCGCAAGCGCGCTAC 121  
DB 1 ATGCCGAGTCGTAGCTGCGCCCAAGACTTGAATGGGCTTCGCAAGCGCGCTAC 60  
QY 122 CAGATCGAAGCGCCGTCGCAAGAGGTGCGCGCGCGCGCTTCATCTGGGACAGTATGC 181  
DB 61 CAGATCGAAGCGCCGTCGCAAGAGGTGCGCGCGCGCGCTTCATCTGGGACAGTATGC 120  
QY 182 CACCTGAGCATTCGCGCACCAACGCGCGCAACGCGCATGTGCTTGCATCACTACAC 241  
DB 121 CACCTGAGCATTCGCGCACCAACGCGCGCAACGCGCATGTGCTTGCATCACTACAC 180  
QY 242 CGCTAGCATAGGACTTGTCTCTTGAACCAATAGCGCGCAAGCGCTACCGCTTCTCC 301  
DB 181 CGCTAGCATAGGACTTGTCTCTTGAACCAATAGCGCGCAAGCGCTACCGCTTCTCC 240  
QY 302 TTGTCTGTGTCGCGGATCATTTCCCTCGCGGAGCGCTGATCCCGTCAACGAGAGGA 361  
DB 241 TTGTCTGTGTCGCGGATCATTTCCCTCGCGGAGCGCTGATCCCGTCAACGAGAGGA 300  
QY 362 ATTGAGTTTACAGCAAACTGATGACGCGCTGTGAGCGCGGATACGCTTGGGTG 421  
DB 301 ATTGAGTTTACAGCAAACTGATGACGCGCTGTGAGCGCGGATACGCTTGGGTG 360  
QY 422 ACTTTGTACACATGGAATCTGCTCAGCGGCTTACAGATGCGATGAGGCTGCTAAC 481  
DB 361 ACTTTGTACACATGGAATCTGCTCAGCGGCTTACAGATGCGATGAGGCTGCTAAC 420  
QY 482 GTGGAAGAGTCCAGCTGACCTTGAAGCGGTATGCGAGTGTGCTTTGAAGCTTTGGG 541  
DB 421 GTGGAAGAGTCCAGCTGACCTTGAAGCGGTATGCGAGTGTGCTTTGAAGCTTTGGG 480  
QY 542 GACCGAGTCCAGAACTGATACCATCAACGAGCCCTGATTCAGGCACTATATGATAT 601  
DB 481 GACCGAGTCCAGAACTGATACCATCAACGAGCCCTGATTCAGGCACTATATGATAT 540  
QY 602 GCCACGCGCAGCAAGCGCGCGCGCAGAGAGATTAACAAGCACTCAACGAGGGAAC 661  
DB 541 GCCACGCGCAGCAAGCGCGCGCGCAGAGAGATTAACAAGCACTCAACGAGGGAAC 600  
QY 662 ACTGCACTAGCGCGTGTGCTGTGGAAGGCCCAAGATCATGAGCATGCGCGCGGTG 721

Db	601	ACTGCCACTGAGGCGGTGGCTCGCTGGAAAGGCCCAATCATGAGCCATGCCCCGCGTG	660
Qy	722	GCCGCTCAACGACGGGACTTTTCCGCCCTCGCAAAAGGGCCAGTGGGCAATTGCTCAAC	781
Db	661	GCGGCTTAACGACGGGACTTTTCCGCCCTCGCAAAAGGGCCAGTGGGCAATTGCTCAAC	720
Qy	782	GGCGACTACTATGAGCCCTGGGACACCAATGAGCCTGGGACAAAGGAGCTGTGAGCGA	841
Db	721	GGGGACTACTATGAGCCCTGGGACACCAATGAGCCTGGGACAAAGGAGCTGTGAGCGA	780
Qy	842	CGAGTGAATTTCACTTGGCTGGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA	901
Db	781	CGAGTGAATTTCACTTGGCTGGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA	840
Qy	902	GAGAGCATGAAGAAGAGCTGGGGCCGAGAGCTTTCCAGCCCTCACTCCCGCGACTTTGCC	961
Db	841	GAGAGCATGAAGAAGAGCTGGGGCCGAGAGCTTTCCAGCCCTCACTCCCGCGACTTTGCC	900
Qy	962	ATCTCTCAATGCCGGAGAGACCGAATTCTTACGGCATGAATTACTACATATCCAGTTCCGG	1022
Db	901	ATCTCTCAATGCCGGAGAGACCGAATTCTTACGGCATGAATTACTACATATCCAGTTCCGG	960
Qy	1022	CCGCAACTGAGAGGTCCGCTCCCGGACGACATCTTGGCGCATATCCATAGACACAG	1081
Db	961	CCGCAACTGAGAGGTCCGCTCCCGGACGACATCTTGGCGCATATCCATAGACACAG	1020
Qy	1082	GAGAAATTAAGACGGGACGCCCGCTTGGCGAGAGAGGCGGCTCGCTGGCTGGCTCCGC	1141
Db	1021	GAGAAATTAAGACGGGACGCCCGCTTGGCGAGAGAGGCGGCTCGCTGGCTGGCTCCGC	1080
Qy	1142	CCGGACATGTTTCCGGAAGCATTTCCGCGGGGTGTACGGCTGTACGGCAAGCCATCTAC	1201
Db	1081	CCGGACATGTTTCCGGAAGCATTTCCGCGGGGTGTACGGCTGTACGGCAAGCCATCTAC	1140
Qy	1202	ATCACCGAGAACGAGATGCCCCGCTCGGAGAGAGAACATGATGGTGGAGAGAGCCGCTC	1261
Db	1141	ATCACCGAGAACGAGATGCCCCGCTCGGAGAGAGAACATGATGGTGGAGAGAGCCGCTC	1200
Qy	1262	AACGACCCCTTCCGATCCGCTACTTTGACTTCGCACTTGGACTCGATTTCCAAAGCCATT	1321
Db	1201	AACGACCCCTTCCGATCCGCTACTTTGACTTCGCACTTGGACTCGATTTCCAAAGCCATT	1260
Qy	1322	ACCCGAGACGGGCTGTGGTCAAGGGGTAACTTTGGCTGGGGCTTGTCTGATTAATTGGAA	1381
Db	1261	ACCCGAGACGGGCTGTGGTCAAGGGGTAACTTTGGCTGGGGCTTGTCTGATTAATTGGAA	1320
Qy	1382	TGGTCAGATGGCTTACGGAACCGAGATTCGACGTTCAAGACTACACCAACCTCTAAG	1441
Db	1321	TGGTCAGATGGCTTACGGAACCGAGATTCGACGTTCAAGACTACACCAACCTCTAAG	1380
Qy	1442	CGCAAGCCCAAGAAATCTGCTCTGTCTCTCAAGACATGTTTGGGCGCGCAAGGGTT	1501
Db	1381	CGCAAGCCCAAGAAATCTGCTCTGTCTCTCTCAAGACATGTTTGGGCGCGCAAGGGTT	1440
Qy	1502	AAAGTGGCGGCATAA 1516	
Db	1441	AAAGTGGCGGCATAA 1455	
RESULT 3			
AAFL14935			
ID AAF14935 standard, cDNA, 735 BP.			
XX AAF14935;			
XX AC			
DT 15-SEP-2003 (revised)			
DT 13-MAR-2001 (first entry)			
XX			
DE Trichoderma reesei EST SEQ ID NO:7458.			
XX			
KW Multiple gene expression; filamentous fungal cell; EST; expressed sequence tag; Fusarium venenatum; Aspergillus niger;			

KW	Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KV	culture condition; environmental stress; spore morphogenesis;
KV	metabolic pathway engineering; catabolic pathway engineering; ss.
XX	
OS	Hypocrea jecorina.
PN	WO200056762-A2.
XX	
PD	28-SEP-2000.
XX	
PF	22-MAR-2000; 2000MO-US007781.
XX	
PR	22-MAR-1999; 99US-00273623.
PA	(NOVO ) NOVO NORDISK BIOTECH INC.
PA	(NOVO ) NOVO NORDISK AS.
PI	Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB,
DR	WPI; 2000-594572/56.
PT	Monitoring differential expression of genes in filamentous fungal cells
PT	uses fluorescence-labeled nucleic acids isolated from the cells and a
PT	substrate of expressed sequence tags.
XX	
PS	Claim 89; Page 3012; 3161pp; English.
XX	
CC	The present invention describes a method for monitoring differential
CC	expression of genes in a first filamentous fungal (PF) cell relative to
CC	expression of the same genes in one or more second filamentous fungal
CC	cells. The method uses fluorescence-labeled nucleic acids isolated from
CC	the PF cells and a substrate of expressed sequence tags (EST). The ESTs
CC	are used in the methods for monitoring differential expression of genes
CC	in a first filamentous fungal (PF) cell relative to expression of the
CC	same genes in one or more second filamentous fungal cells. Monitoring the
CC	global expression of genes from PF cells allows the production potential
CC	of the microorganisms to be improved. New genes may be discovered,
CC	possible functions of unknown open reading frames can be identified and
CC	gene copy number variation and stability can be monitored. The expression
CC	of genes can be used to study how PF cells adapt to changes in culture
CC	conditions, environmental stress, spore morphogenesis, recombination,
CC	metabolic or catabolic pathway engineering. Using ESTs provides several
CC	advantages over genomic or random cDNA clones including elimination of
CC	redundancy as one spot on an array equals one gene or open reading frame,
CC	and organization of the microarrays based on function of the gene
CC	products to facilitate analysis of the results. AAF07478 to AAF11247
CC	represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents
CC	ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from
CC	Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from
CC	Trichoderma reesei, which are all specifically claimed in the present
CC	invention. (Updated on 15-SEP-2003 to standardise OS field)
XX	
SQ	Sequence 735 BP; 168 A; 210 C; 205 G; 148 T; 0 U; 4 Other;
	Query Match 30.0%; Score 598; DB 3; Length 735;
	Best Local Similarity 97.0%; Pred. No. 1e-137;
	Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4
OY	319 CATTCCTCGGCGGAGGCTGTATCCCGTCAACGAAGAAGAAATTGATTTTACAGCAA 378
Db	1 CAATTCCTCGGCGGAGGCTGTATCCCGTCAACGAAGAAGAAATTGATTTTACAGCAA 60
OY	379 ACTGATTGAACGCCGTTTGAGGGGGGTATCAGCCCTTGCGTAGACTTTTACCACCTGGGA 438
Db	61 ACTGATTGAACGCCCTTGTTGAGAGCGGGGTATCAGCCCTTGCGTAGACTTTTACCACCTGGGA 120
OY	439 TCTGCCTCAGGCGCTTACAGATCGCTATGTAGAGGCTGGCTCAACGTGGAAGAGTCCAGCT 498
Db	121 TCTGCCTCAGGCGCTTACCATGATGAGGCTGGCTCAACGTGGAAGAGTCCAGCT 180
OY	499 GGACTTTGAGCGGTAATGCGAGGTTTGCTTTGAACGTTTTGGGAGCCGAGTCCAGAAC TG 558
Db	181 GGACTTTGAGCGGTAATGCGAGGTTTGCTTTGAACGTTTTGGGAGCCGAGTCCAGAAC CTG 240

QY 559 GATCACCATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACGCAACGC 618  
 DB 241 GATCACCATCAAGMCCCTGATTGAGCCATCTATGATATGCAACCGGACGCAACGC 300  
 QY 619 CCGGGGAGAGAGAGCATTAACAAGCACTCCACCGAGGGGCAACCTGCCACTGAGCCGTG 678  
 DB 301 CCGGGGAGAGAGAGCATTAACAAGCACTCCACCGAGGGGCAACCTGCCACTGAGCCGTG 360  
 QY 679 GCTGCTGGAAGAGCCCAATCATGAGCCATGCCCGCGCGTGGCGCTCTACAGCAGGA 738  
 DB 361 GCTGCTGGAAGAGCCCAATCATGAGCCATGCCCGCGCGTGGCGCTCTACAGCAGGA 420  
 QY 739 CTTTCCGCCCCCGCAAAAGGGCCAGATGCGCATCTGCTCAAGGGGAGTACTATGAGCC 798  
 DB 421 CTTTCCGCCCCCGCAAAAGGGCCAGATGCGCATCTGCTCAAGGGGAGTACTATGAGCC 480  
 QY 799 CTGGGACAGCATGAGCTCGGGGCAAGAGAGCTGTAGAGGACGATGGAATTTCAAT 858  
 DB 481 CTGGGACAGCATGAGCTCGGGGCAAGAGAGCTGTAGAGGACGATGGAATTTCAAT 540  
 QY 859 TGCGTGTGTTGCCATCTCTTTTGAAGAGAGCTA-TCCAGAGCATGAAGAGC 917  
 DB 541 TGCGTGTGTTGCCATCTCTTTTGAAGAGAGCTA-TCCAGAGCATGAAGAGC 600  
 QY 918 AGCT--GGGGAGAGGCTTCAGCCCTCAC-TCCGGGAGCTTT--GCCATCTCATATGC 972  
 DB 601 AACTTGGGGGAGAGGCTTCACCTTACCTTCCCGGAGCTTTTGCATCTTAAATGC 660  
 QY 973 CGGAGAGAGC 982  
 DB 661 CGGAGAGAGC 670  
 RESULT 4  
 AAF14882  
 ID AAF14882 standard; cDNA; 1933 BP.  
 XX AAF14882;  
 AC AAF14882;  
 XX 15-SEP-2003 (revised)  
 DT 13-MAR-2001 (first entry)  
 XX  
 DE Trichoderma reesei EST SEQ ID NO:7405.  
 XX  
 KW Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Hypocrea jecorina.  
 OS  
 PN WO200056762-A2.  
 XX  
 PD 28-SEP-2000.  
 XX  
 PF 22-MAR-2000; 2000WO-US007781.  
 XX  
 PR 22-MAR-1999; 99US-00273623.  
 XX  
 PA (NOVO) NOVO NORDISK BIOTECH INC.  
 PA (NOVO) NOVO NORDISK AS.  
 XX  
 PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;  
 DR WPI; 2000-594572/56.  
 XX  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags.  
 XX  
 PS Claim 89; Page 2989; 3161p; English.

XX The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring the  
 CC global expression of genes from FF cells allows the production potential  
 CC of the microorganisms to be improved. New genes may be discovered,  
 CC possible functions of unknown open reading frames can be identified and  
 CC gene copy number variation and stability can be monitored. The expression  
 CC of genes can be used to study how FF cells adapt to changes in culture  
 CC conditions, environmental stress, spore morphogenesis, recombination,  
 CC metabolic or catabolic pathway engineering. Using ESTs provides several  
 CC advantages over genomic or random cDNA clones including elimination of  
 CC redundancy as one spot on an array equals one gene or open reading frame,  
 CC and organization of the microarrays based on function of the gene  
 CC products to facilitate analysis of the results. AAF07478 to AAF11247  
 CC represents ESTs from Fusarium venenatum; AAF1248 to AAF1853 represents  
 CC ESTs from Aspergillus niger; AAF1854 to AAF1878 represents ESTs from  
 CC Aspergillus oryzae; and AAF14879 to AAF1537 represents ESTs from  
 CC Trichoderma reesei, which are all specifically claimed in the present  
 CC invention. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1933 BP; 467 A; 536 C; 568 G; 343 T; 0 U; 19 Other;  
 Query Match 13.4%; Score 266.6; DB 3; Length 1933;  
 Best Local Similarity 56.2%; Pred. No. 1.5e-55;  
 Matches 770; Conservative 1; Mismatches 540; Indels 59; Gaps 13;  
 QY 81 TGCCCAAGCATTTGATGGGCTTGGCAAGCGCCGCTTCCAGATGAGAGGCGCTCA 140  
 DB 250 TGCCCAAGCATTTGATGGGCTTGGCAAGCGCCGCTTCCAGATGAGAGGCGCTCG 309  
 QY 141 AAGAAGGTGGCCGCGCCGCTTCCAT--CTGGGACAGCTACTGC-----CACTGAGGCCA 193  
 DB 310 ACCAGAGAGCGCCGCGCCGCTTCCAGATGAGAGGCGCTTCCAGATGAGAGGCGCCG 369  
 QY 194 TGCGGACCAAGCGCGCCGCTTCCAGATGAGAGGCGCTTCCAGATGAGAGGCGCTT 252  
 DB 370 TGCCGAGCGGCTTCCAGATGAGAGGCGCGCTTCCAGATGAGAGGCGCGCTTCCAG 429  
 QY 253 GACATTTGATCTCTTGAACCAAGTACGCGGCAAGGCTTCCGCTTCTTGTGCTG 312  
 DB 430 GACATTTGATCTCTTGAACCAAGTACGCGGCAAGGCTTCCGCTTCTTGTGCTG 489  
 QY 313 GCGGATCATTCCTCGCGCGGAGGCTTCCGCTTCCAGATGAGAGGCGCTTCCAGAT 372  
 DB 490 GCGGATCATTCCTCGCGCGGAGGCTTCCGCTTCCAGATGAGAGGCGCTTCCAGAT 549  
 QY 373 CAGCAACCTGATGAGCGCCCTGTTGAGGCGGAGTATCAGCCTTGGTGACTTTTACCA 432  
 DB 550 CAGCAACCTGATGAGCGCGCTTCCGCTTCCAGATGAGAGGCGCTTCCAGATGAGAG 609  
 QY 433 CTGGGATCTGCTCAAGCGCTTCCAGATGAGAGGCGCTTCCAGATGAGAGGCGCT 492  
 DB 610 CTGGGATCTGCTCAAGCGCTTCCAGATGAGAGGCGCTTCCAGATGAGAGGCGCT 669  
 QY 493 CAGCTGATCTTGAAGGCGGATGAGAGGCTTCCGCTTCCAGATGAGAGGCGCTTCCAG 552  
 DB 670 CAGCTGATCTTGAAGGCGGATGAGAGGCTTCCGCTTCCAGATGAGAGGCGCTTCCAG 728  
 QY 553 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 612  
 DB 729 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 788  
 QY 613 CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 672  
 DB 789 CTTGCGCCCGCGCGCGGAGCA-----CTTGGGA 818  
 QY 673 GCGTGGCTGCTGGAAGAGCCAGATCATGAGCATGCGCGCGCGCTGCTACAG 732

```

Db      819 GCGGTGACCGTGGCCACAACATCTGTCGCCACGCGCGCCCTCAAGGCGTACC 878
      733 CAGGAGCTT---TGGCCCTTCGCAAAAGGCGAGATGCGATCTCGTCAACGGCGACTA 789
      879 CGAGACCTTCAAGCCCGCCGACGGCGAGCGGCAGATGGCATGCGTCTTCAAGCGGACTT 938
      790 CTATGAGCCCTTGGGACGACATAGAGCTCGGACCAAGAGGCTGCTGAGCGAGCGATGGA 849
      939 CACTACCTCCCTGGAGACCGCCCGGACCGGCGGACAGAGGCGGCGGCGGCGCTCGA 998
      850 ATTTCACATTGGCTGGTGTTCGCAATCCATCTTCTTGAAGAGAGCATATCCAGAGCAT 909
      999 GTTCTTCAAGCGCTGGTTCGCGAGATCCCATCTTGGGCGACTAACCGGCGTCA--TGC 1057
      910 GAAGAGACGCTGGGCGGAGAGCTTCCAGCCCTCACTCCCGGAGCTTGTGCTCTCAA 969
      1058 GCAAGCAAGCTGGGCGGACCGGCTGCGCACTTTACGCCGAGGA--AGCGCGCTCTGT 1115
      970 TGGCGAGAGACCGACTTCTACGCGATGATTAATTAACATCCAGTTGCGCGCGACCT 1029
      1116 CCACGGCTTCAACGACTTTACGGGATGAAACCACTACAGTCCAACTACATCCGACCGA 1175
      1030 AGACGCTCCCTGCGCGGAGAGGACTATCTGGGCGGCAATCATAGACGACGAGATTA 1089
      1176 AGCTCG--CCGCTTGGCGGACGACCGCTGGCAAGCTGCACTGCTTCCACCAACA 1233
      1090 GACGCGGACCGCCGTTGGCGAGAGAGCGGCTCGCTGCGTGGCTCT--GGCCGAGC 1147
      1234 GCAAGGCAATGTCATCGGCGCCCGGAGA--CGAGTCCCTCGGCTGGGCGCTTGTGCGCG 1292
      1148 ATGTTCCGGAAGCATCTGCGCGGCTGTAAGGCTCTACGCGCAAGCCCATCTACATCAC 1207
      1293 GATTTCGCCGANTTCTNGTGTGTGATTCAGCAAGANGACNGGTACCCGCCATTNTACGT 1352
      1208 GAGAACGATGCGCGCTGCGCTGAGAGAGAAATGACGTGC-----GAGGAGCGCGT 1260
      1353 GACGGANACGAGNACGAGCATTTAAGGGCGAGAGGACTTGGCAAGAGAAAGAAATNT 1412
      1261 CAACGACCCCTTCCGCAATCCGCTACTTTGACTCGCACTTGGACTGAT--TTCCAGGCGC 1318
      1413 CGAAGTGACTTCAAGGTCAATTAACAAGTACATCCGTCATGGGTACCCGCG 1472
      1319 ATTACCCAGAGAGCGCTGCTCTCAAGGCTACTTGGTGGGCGTTCGATTAATTTG 1378
      1473 TGGAGCTGACGCGGCTCAAAAGTCAAAAGGCTACTTGGCTGCTCATGGAACAATT 1532
      1379 GAATGTCAGATGCTACGAGACCGAGATTGCGCTGACGTTCAAGACTTA 1428
      1533 GAGGGGCGGACGCTACGTAAGAAAGTTTGGGGTTACTATGGGGAATTA 1582

RESULT 5
ADA70881
ID      ADA70881 standard; DNA; 1404 BP.
AC      ADA70881;
DT      20-NOV-2003 (first entry)
XX      Rice gene, SEQ ID 4204.
XX      Plant; bacterial infection; fungal infection; viral infection; rice;
KW      gene; ds.
OS      Oryza sativa.
XX      PN      MO2003000898-A1.
XX      PD      03-JAN-2003.
XX      PF      22-JUN-2001; 2001WO-1B001105.
XX

```

```

PR      22-JUN-2001; 2001WO-1B001105.
PA      (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX      Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
PI      Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX      WPI; 2003-175290/17.
XX      DR
XX      PT      Identifying at least one gene involved in plant resistance or response to
PT      pathogenic infection for conferring resistance or tolerance to a plant to
PT      bacterial, fungal or viral infection by determining or detecting plant
PT      gene expression.
XX      PS      Claim 6; SEQ ID NO 4204; 899bp; English.
XX      SS
XX      CC      The present invention relates to a method (M1) for identifying genes
XX      involved in plant resistance or response to pathogenic infection. M1
XX      comprises identifying a gene whose expression is significantly altered in
XX      the incompatible interaction of plant gene expression relative to
XX      expression of the gene in an uninfected plant, in a mutant plant that
XX      does not express a gene associated with response to pathogenic infection,
XX      or in a corresponding incompatible or compatible interaction. (M1) is
XX      useful for conferring resistance to resistance or tolerance to a plant to
XX      bacterial, fungal or viral infection. The present sequence was used to
XX      illustrate the invention.
SQ      Sequence 1404 BP; 356 A; 385 C; 396 G; 266 T; 0 U; 1 Other;
Query Match      12.4%; Score 246; DB 7; Length 1404;
Best Local Similarity 54.5%; Pred. No. 1.7e-50;
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

      123 AGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCCGCTCATCTGAGACCACTATGCC 182
      86 AGTACGAGGAGGCTGTGAAGAGAGAGCGAGAGGACGACATCTGGGACACGTTGGCGC 145
      183 ACTGAGACCATCGCGCACCAACGCGCGCAAGCGCGATGTGGCTTGGCATCACTACACC 242
      146 ACACCTTTGGAAAGATCACCGACTTCAGCAATGCTGATGTGGACGTTGATCAGTACACC 205
      243 GCTACGATGAGACTTGTGATCTCTTACCAAGTACAGGCGGAAAGGCGCTTCTCTCT 302
      206 GTTTCGAGAGGAGATACACTCATGCGCAAGCATGGAGATGATGCTATCGGTTCTCGA 265
      303 TGTGCTGCTGCGGATCATTCCTCCGCGCGGAGGCTGATGCCGTCACAGAGAGGAAA 362
      266 TAGCATGTCAAGATCTACCCAAATGTTG-----TTGGTCAAGTCAATCAAGCTGTA 319
      363 TTGAGTTTACAGCAAACTGATTGACGCGCTGTTGAGGCGGGTATCAAGCTTTGGTGA 422
      320 TCGACACTACACAAAGCTGATGATGATCTTACGCAAAAGAAATTCAGCATATATGTA 379
      423 CTGTGTACCACTGGATGCTGCTCAGGCGCTTTCAGATCGCTATGAGAGGCTGCTCAACG 482
      380 CACTTACCACTGGAGACTTCCCGAGGCGCTTGAAGACATGAAGAGGCTGGCTTGACA 439
      483 TGAAGAGTCCAGCTGACTTTGAGCGGTAATGCGAGTGTGCTTGAAGCTTTTGGGG 542
      440 GG---CAATAGTGGAGCATTTGCGGGGTACGCGGAGACGTCGCTTCAAGGAGTGGGG 496
      543 ACCGATCCAGAACTGTGATCACTACACGACGCTGATTCAGGCACTATGATGATATG 602
      497 ACAGGTGAAAGCACTGATGATCAAGCTCAAGCGGACACAGGTGGCCATTCAGAGGCTACG 556
      603 CCACGCGGACGACGCGCGCGGAGAGCAGACATTA-----ACAAGCATCTCAACGAGG 656
      557 ACGACGGGCTCAGGCGCGCGCGCGCTGCTCGTGTCTCACTCACTCAAGAGCGCG 616
      657 GCAACATGCACTGAGCGGTGCTGCTGGAAGAGCCCAATCATGAGCATGCGCGCG 716
      617 GCAACTCGGCGGACGAGCCCTAAGCTGCTGCGCCACCACTTATCTCGCCACGCGCGCG 676

```

QY 717 CCGTGGCGCTTACAGAGGAGCTTTCGCCCTTCGCAAAAAGGCGCAGATCGCTCGC 776  
 DB 677 CCGGCACATCTACAGACAAATAACAGGCGACGCAAGGCGACCTTGGATACCT 736  
 QY 777 TCACGGCGACTACTATAGCCCTTGGAGACAGCAATGAGCCTCGGACAAAGAGCTGCTG 836  
 DB 737 TCGACGTGATGTGGTTCGAGCCGATGTCACACACAGAT---CGACATGAGGCGGCGCA 793  
 QY 837 AGGAGCGGATGGAATTTACATATGGCTGCTGCTTCCATTTCCATTTTGAAGAGACT 896  
 DB 794 AGAGAGCGAGAGGTTTACATAGATGTTTGTGATCCGTTCTT---CGGCACT 850  
 QY 897 ATTCAGAGACATGAGAGACAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACT 956  
 DB 851 ACCCGGAGAGATGAGGCGGAGCTGGGAGAGAGCTGCCAGTTTCACGCGGATGAGG 910  
 QY 957 TTGCCATCTCAATGCCGAGAGACCGACTTCTAAGGCAATTAATTAACATCCAGT 1016  
 DB 911 CCGCGCTGCTCAAG---GGGCGCTGGATTTCGTGCGCATTAACCACTACACCACTACT 967  
 QY 1017 TCGCGCGCCACTAGAGGCTCCCTCCCGAGA-----CGGACTATCTCGCGCCACTCA 1071  
 DB 968 ACAGAGAGCAACACACCAACCAATCATCGGAGATTCCTCAACACACCTTGGCAGACA 1027  
 QY 1072 TGAGCAGCAGAGAAAT-----AAGAGCGCAGCGCCGTTGGCGAGAGAGCGGCTCG 1124  
 DB 1028 CCGGACCGCTGAGCTCCCATTCAGAAATGGGAAGCCAAATTGGAGATAGGCGAAATTGCA 1087  
 QY 1125 CTTGGCTGCGCTCTGCGGAGACATGTTCCGAGAGCATCTCGCCGGGTGTACGCGCTGT 1184  
 DB 1088 TATGGCTGTACATTTGTCGCCGAGAGATGAGAGCTGATGAATATGTCAAGGAAGGT 1147  
 QY 1185 ---ACGCGAAGCCCATTAATCATCCGAGAACCGATCCCGCTCGAGAGAGAACCA 1241  
 DB 1148 ACAAGAGCCCAACAGTGTACATCACTGAAAACGGGATGATGACAGACAAACCGTTCA 1207  
 QY 1242 TGACGTGCGAGAGGCGGTCAACGACCCCTTCGCGATCCGTAATTTGATCGCACTTGG 1301  
 DB 1208 TTTTCATCAAGAGAGCGCTTCAAGAGACAGCAAGAGATCAATATACCAATGACTACTCA 1267  
 QY 1302 ACTGATTTTCAAGGCCATTACCCAGAGACGCGCTGCTCAAGGGGTACTTTGCGTGG 1361  
 DB 1268 CCAATCTGCGCTGCTTCATCAAGAGAGCGGCTGCGACGTAACGGGTACTTGGCGGTG 1327  
 QY 1362 CGTTGCTCGAATCTTGGAA-TGTCAGATGGCTACGAGACCCGATTCGCGCTACCTTC 1420  
 DB 1328 CTCTGCTGAGCAACTGGGAATGGGCGGCGGATATCTCTGAGATTGCGGCTCTACTTC 1387  
 QY 1421 ACAGACTACA 1430  
 DB 1388 GTGGACTACA 1397

RESULT 6  
 ADA69524  
 ID ADA69524 standard; DNA; 1503 BP.

AC ADA69524;  
 XX 20-NOV-2003 (first entry)  
 DT XX  
 DE Rice gene, SEQ ID 2847.  
 KW plant; bacterial infection; fungal infection; viral infection; rice;  
 gene; db.  
 XX  
 OS Oryza sativa.  
 XX  
 PN W02003000898-A1.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001MO-IB001105.

XX  
 PR 22-JUN-2001; 2001MO-IB001105.  
 XX  
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
 XX  
 DR WPI; 2003-175290/17.  
 XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 PS Claim 6; SEQ ID NO 2847; 899bp; English.  
 XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 SQ Sequence 1503 BP; 380 A; 420 C; 379 G; 324 T; 0 U; 0 Other;  
 Query Match 10.8%; Score 214.6; DB 7; Length 1503;  
 Best Local Similarity 51.0%; Pred. No. 1e-42;  
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;  
 QY 81 TGCCTCAAGCACTTGAATGAGGCTTCCGACAGGCGCCCTCAACGATGAAAGCGCGCTCA 140  
 DB 95 TCCCAAGAGCACTTCACTTTCGCGACAGGTTCAAGCACTTATCAGTATGAGGCGCTTACA 154  
 QY 141 AAGAAGTGGCGCGCGCGCGCTCCATCTGAGAGACAGTACCTCCAGTGGAGCCATCGGCA 200  
 DB 155 AAGAAGGCGCGCGCGCGCGCTTCAAGCTGCGGACACTTCACTCACTTCCAGTAAATTT 214  
 QY 201 CCAAGCGCGCGCGCGCGATGTGGCTTGCATCAACAGCGCTAGAGAGACTTTG 260  
 DB 215 TAAACATATGTAACCGCGATGTGGCAATGACTTTCATACCGATACAGAGAGATGTGA 274  
 QY 261 ATCTTGAACCAATGACGCGCAAGCGCTTCTCTTGTGCTGCTGCGGATCA 320  
 DB 275 GCTCTGGAAGGACATGAACATGATGCTTCCGTTCTCATTTGCTGAGACAGGATTC 334  
 QY 321 TTCCCTCGCGCGCGAGCTGATCCGCTCAACGAGAGGAAATTTAGTTTACGCAAC 380  
 DB 335 TGCCTAATGATCTTGAAGTGAAGAAATAAACAAGAGGATTTGCTTCAACAGCT 394  
 QY 381 TGAATGACCGCTTGTGAGGCGGAGTATCAAGCTTGGGATCTTTGTACACTTGGATC 440  
 DB 395 TGAATGATGATGATGATGCAAAAGGATGATCCATTTGTGATCTTCACTTGGAGCA 454  
 QY 441 TGCCTGAGCGCTTCAAGATCGCTATGAGAGCTGCTCAACGTGGAAGAGTCAAGTGG 500  
 DB 455 CCCCTTGGCTTGAAGAGCAATTAAGAGAGATTTCTCA---GTGAAGACATAGTGAAG 511  
 QY 501 ACTTTGAGCGGTATGAGAGTGTGCTTTTGAAGCTTTTGGGAGCCGAGTCAAGACTGGA 560  
 DB 512 AATAGTGAACCTTGGGAGAGTGTGCTTTCGAGATTCGCGACCGTGTCAATTAATCTGGA 571  
 QY 561 TCACATCAAGAGAGCTGATTCAGAGCATATGAGATATGAGCAAGGAGCAAGCGCC 620  
 DB 572 CCAATTTAATGAGCACTTCAATATAGCCCTTACGCTACAGGAGGAGGTGTTTGGCG 631  
 QY 621 CCGGAGAGAGAGCA-----TTAACAAGCACTTCAACGAGGCAACACTTGCACAGAGC 674  
 DB 632 CCGGAGAGAGTCTCTTCAATATGTTTCAAGATGATGCGGCGGTGACTCAGTCCGAGAGC 691

QY 675 CGTGGCTCGGTGAAAAGGCCCAATCATGAGCCATGSCCGCGGTGGCCGTCTACAGCA 734  
 DB 692 CCGACTCGGTGGGACACCAATCCACTCTCCACGCTGCCCTGTCCAGCTTTACCGCA 751  
 QY 735 GGGACTTTGGCCCTTCGCAAAAAGGCGAGATCGGCACTCTGCTCAACGGCGACTACTATG 794  
 DB 752 CCAGATGACGACCAACACAGAAAGGACAGATCGGATGGTGGTGTCAACCCACTGGTTCG 811  
 QY 795 AGCCCTGGAGACAGATGAGCCTCGGGACAAGAGAGCTGTGAGGAGCGAATGGAATTC 854  
 DB 812 TGGCGGTGACCACTTCCAGCGT---GACCGTGGCGCTGTGCAAGGAGGAGCTTGAAGCTTCA 868  
 QY 855 ACATTGGCTGTTTGCCCAATCCCATCTTTGAGAAAGGACTATCAGAGAGCATGAGA 914  
 DB 869 TCTATGGGTGGTTCATGAGACCTTA---TCGTGATGTGTGACTACCAAGGACCAATGAGAG 925  
 QY 915 AGCAGCTGGGGCGAGAGGCTTCCAGCCCTGACTCCCGGAGCTTTGCCATCTCTCAATGCGG 974  
 DB 926 GTTGGCTCGGCAATCGGCTGCGCAGAGTTCAAGCTTGAACAGTGGCGGATGGTGAAGGGCT 985  
 QY 975 GAGAGCCGACTTCTACGGCATGAATTACTACATCCCATGTTGGCGGCAACCTAGACG 1034  
 DB 986 CCTA---CGACTTCTACGAGTAACTTACTACACCACTTACGCTTAAGATATACCGC 1042  
 QY 1035 GTCCGCTCCCGA-----GACGACTATCTTGCGCGCATTCATGAGACGAGAGATA 1088  
 DB 1043 CGCCTTACTCCAAACGAGTATCTTACGACTGACCTGACACCGCGCAACACACCGGCTTCC 1102  
 QY 1089 AGGACGGAGAGCGCGGTGGCGAGAGAGAGAGCGGCTGCGCTGGCTGCTGCTGCGGAGCA 1148  
 DB 1103 GTATGCGAAACCCATCGGTCCACAGGAATTTACACCCATCTTCAACTAACCTTCCAG 1162  
 QY 1149 TGTTCGGAAGATCTCGCCCGGAGTACGCGCTGTGACGCAAGGCCATCTTACATCAACG 1208  
 DB 1163 GTCTCGTAGAGCTCTCTCTTACACCAAGAGAGATTAACAACCCGA---CAATCTATG 1219  
 QY 1209 AGAAGGATGCGCGCTGCGAGAGAGAGAAACATGACCTGCGAGAGGCGCTCAACGACC 1268  
 DB 1220 TTACAGAAACCGGACATCATGAGGTTAACACAGCACCTGCCAGAGGCGCTCAAGATG 1279  
 QY 1269 CTTTCGCGATCGGTACTTTGACTGCGCACTTGGAATGATTTCAAGGCCATTTACCAAG 1328  
 DB 1280 GACACAGATGAGATTCACATCAAAAGCACCTGCGAGTTGTTCAACCAAGCGCAT---CAAGA 1336  
 QY 1329 ACGCGCTGCTGCAAGGGGTACTTTGCGTGGGCGTGTGCTGATTAATCTTGGATGTGCTAG 1388  
 DB 1337 ATGGGTGAGACGTGAAGGGTACTTCAAGTGAATTCATGGAATGCTGAGTGGGGTG 1396  
 QY 1389 ATGGCTAGGACCAAGATTTGGCGCTGACGTTCAAGACTACACCACTTCAAGCGCACGC 1448  
 DB 1397 ACGGTACTTGAACAGGTTGGGCTCTCATCTACGTCGACCGCAAGACCTCAAGGCTAAC 1456  
 QY 1449 CCAAGAAGTCTGCTGCTCTCAAGGACAT 1479  
 DB 1457 GCAAGAGTCCAGCTACTGATCGAAGACTT 1487

RESULT 7  
 ABZ53130  
 ID ABZ53130 standard; cdna; 575 BP.

XX AC ABZ53130;

DT 28-MAR-2003 (first entry)

DE Aspergillus oryzae polynucleotide SEQ ID NO 2243.

KM Aspergillus oryzae; fermentation; fungus; industrial; BST;

XX expressed sequence tag; gene; ss.

OS Aspergillus oryzae.

XX

PN W0200279476-A1.  
 XX 10-OCT-2002.  
 PD  
 XX  
 PF 22-MAR-2002; 2002WO-IB000890.  
 PR 30-MAR-2001; 2001JP-00098371.  
 XX  
 PA (NABD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
 PA (NABE-) NAT RES INST BREWING.  
 PA (NORQ) NAT FOOD RES INST MIN AGRIC.  
 PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horinouchi H;  
 PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;  
 DR WPI; 2003-046817/04.  
 XX  
 XX  
 PT  
 PT  
 PS  
 CC Claim 1; SEQ ID NO 2243; 48bp + Sequence Listing; Japanese.  
 CC The invention relates to a polynucleotide having any of 6006 specific  
 CC sequences (ABZ50888-ABZ56893), which are expressed by a fungus under  
 CC specific culture conditions including one or more of eutrophic,  
 CC oligotrophic, solid, early germination, alkaline, high temperature, low  
 CC temperature or maltose culture or polynucleotides stringently hybridizing  
 CC to these sequences. The polynucleotides are useful for monitoring the  
 CC progress of fermentation and the growth conditions of a fungus,  
 CC especially of Aspergillus oryzae which is widely used in industrial  
 CC fermentation. Also monitoring for fungal contamination. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ffp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 575 BP; 143 A; 154 C; 134 G; 144 T; 0 U; 0 Other;

Query Match 10.3%; Score 205.4; DB 7; Length 575;  
 Best Local Similarity 63.0%; Pred. No. 1,3e-40;  
 Matches 317; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 78 CTCTGCCCAACGACCTTGAATGGGGCTTGGCAACGGCGGCTTACAGATGAAAGGCGCG 137  
 DB 71 CTCTCCACCGGACCTTCTATGGGGATTTGCAACGCGAGTTACAGATTTGAAGGCGCG 130  
 QY 138 TCAAGAAGTGGCGCGGCGCGGCTGATCTGAGGACACGTAACCTGAGGACATGCGC 197  
 DB 131 TAAATGAGACGAGACGAGTCCATTCATCTGAGATCAATTTGCAAGATCCCCGGTAATA 190  
 QY 198 GCAACCAAGCGCGCAACGCGAGTGTGCTTGGATCTACTACCAAGCTTACATGAGAGACT 257  
 DB 191 TTGCTGAAGCGCGCAATGGCGATGTGGCTTGGACATCAATACATGACACACAGAGATA 250  
 QY 258 TTGATCTCTTGAACAAATGACGCGCAAGGCGCTTACCGCTTCTCTTGTGTGTCGCGA 317  
 DB 251 TTGCTTGTCTCAAGGCTTGGCTGCGGCAAGGCAATACGTTTCTCACTCTTGTGCCCA 310  
 QY 318 TCAATTCCTCGGCGGAGCGGATGATCCGTAAGAGAGGAGATGATTTACAGA 377  
 DB 311 TTAATTCACCTCGGTGGCGGAAACGACCCCATTAATGAGAAAGGCTTGCATAATTAATA 370  
 QY 378 AACTGATGACGCGCTGTTGAGGCGGGTATCACCGCTTGGGTGACTTTGTAACACTGGG 437  
 DB 371 AGTTGCTGACGACCTGATGCTGTGCTGATCATCTCTGTTAATTTGTTGACTGGG 430  
 QY 438 ATCTGCTTACGCGGCTTACGATCGCTATGAGAGCTGTGCTCAACGTGAAGAGTCCAGC 497  
 DB 431 ATCTTCGCGATAGCTTCGACAAACGCTACGCGGCTCTCTCATTAATGAAGAGTTCGTCG 490  
 QY 498 TGAAGCTTGAAGCGGATGAGAGGTTGTGCTTTGAAGCTTTTGGGAGCCGAGTCCAGACT 557  
 DB 491 CAGACTTGGCCCATACGACGACGATAGTTTCAAGGCTTTGGGCTCGAAAGTAAAGCAATT 550

QY 558 GGATCACCATCAAGCAGCCCTGG 580  
|||||  
Db 551 GGATCAGCTTCAAGCAGCATGG 573

RESULT 8  
AB213477

ID AB213477 standard; DNA; 1734 BP.

AC AB213477;

DT 21-JAN-2003 (first entry)

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1282.

KM Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

OS Arabidopsis thaliana.

PN MO200216655-A2.

PD 28-FEB-2002.

PF 24-AUG-2001; 2001WO-US026685.

PR 24-AUG-2000; 2000US-0227866P.

PR 26-JAN-2001; 2001US-0264647P.

PR 22-JUN-2001; 2001US-0300111P.

PA (SCRI ) SCRIpps RES INST.

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

PI Harper JF, Krops J, Wang X, Zhu T;

DR WPI; 2002-304127/34.

PT Identifying a stress condition to which a plant cell has been exposed and  
XX producing plants with increased tolerance to these abiotic stresses.

PS Claim 144; SEQ ID NO 1282; 577bp + Sequence listing; English.

CC The invention relates to identifying a stress condition to which a plant  
XX cell has been exposed, comprising: (a) contacting nucleic acid

CC representative of expressed polynucleotides in the plant cell with an  
XX array or probes representative of the plant cell genome; and (b)

CC detecting a profile of expressed polynucleotides in the plant cell  
XX characteristic of a stress response. The method is useful in the

CC production of transgenic plants, cells and seeds and in producing plants  
XX with increased tolerance to abiotic stress. The present sequence is that

CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used  
XX in methods of the invention. Note: The sequence data for this patent is

CC not represented in the printed specification but is based on sequence  
XX information supplied to Derwent by the European Patent Office

XX Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 9.2%; Score 182.8; DB: 6; Length 1734;

Best Local Similarity 49.2%; Pred. No. 8.1e-35;

Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

QY 79 TCTGCCCAAGCAGCTTGAATGGGCTTCGCAACGGCGCCCTACAGATCGAAGGCGCGT 138  
|||||

Db 90 TTTTCCGATGATTTTCACTCTTGGACAGCTCGCTCGCGCTTTCAGTACGAAGGTGCAAC 149  
|||||

QY 139 CAAGAAAGTGGCGCGCGCGCTTCATCTGGACACGTACTGCCACTGAGCCATCGCG 198  
|||||

Db 150 AAGGAAGGTGGCAAGTCTCCAACTATATGGATCACTTCAGCTCCAGTATCCAGAAG 209  
|||||

QY 199 CACCAAGCGGCCCAACGGCATGTGGCTTGGATCACTACACCGCATCAATGAGGACTT 258  
|||||

Db 210 GACCAAAATGATATACAGATGTAGCAATTGATTTTATCATGTTTCAAGAGTGAAT 269  
|||||

QY 259 TGATCTCTGACCAAGTACGGCGCAAGGCTTACCGCTTCTTGTGTGGTGGCGGAT 318  
|||||

Db 270 AAATATGATGAAGAGCTAAACATGACGCTTTCCGATTTTCAATCTCGGTGTCMAATT 329  
|||||

QY 319 CATTCCTCCGCGGCGAGCGCTGATCCCGTCAACGAGAGGAAATTAGTTTACAGCAA 378  
|||||

Db 330 AATACCAATGAAAGCTAAAGATGGAATGAAACAAAGAGGTGTACAAATTTCAAGAGA 389  
|||||

QY 379 ACTGATTTGACGCGCTGTGAGGCGGGGTATACGCGCTTGGGTGACTTTTATCACTGGGA 438  
|||||

Db 390 TCTCATGACGAACTTCTTGTCTATATGACATACAACTTGCATGACGCTTATCATTTGGGA 449  
|||||

QY 439 TCTGCTCAGGCGCTTCAAGATGCTATGAGAGCGCTGCTCAAGCTGAGAGGTCCAGCT 498  
|||||

Db 450 CCAACCCAAATCTTTGGAGAGCAAAATGTGTGCTTTCTTAAGCCCTTAAATTCGTAGA --- 506  
|||||

QY 499 GCACTTGAACCGGATGCGGAGTGTGCTTTGAAACGTTTGGGACCGGATCCGAATCG 558  
|||||

Db 507 AGATTTTCAGATTTTGCAGAAATTTGTTTGAAGAGTTTGAAGATTAAGTTAAGATGTG 566  
|||||

QY 559 GATCACCATCAAGCAGCCTGATTCAGGCACTATGATATATGCCACCGGACAGCAAGC 618  
|||||

Db 567 GACAAATCAAGCACTTATATATGATGTTGGCGGTTATGATCAAGTAAAGAGC 626  
|||||

QY 619 CCGGGGAGGAGC-----AGCATTTAACAAGCACTCCACGGAGGCAACCTGCACTGA 672  
|||||

Db 627 GGTGAGAGATGCTCAAAATGGGTAAACGAAAGGTGCAGGCTGAGATTCAGATACGA 686  
|||||

QY 673 GCGTGGCTCGCTGGAAGAGCCAGATCATGAGCCATGCCCCGCGCTGCGCTACAG 732  
|||||

Db 687 GCTTACATTTGTTTCAATCAATCACTCTTGTGCCAAGCGCTGCAATGGAAGAAATTCG 746  
|||||

QY 733 CAGGAACTTTGCGCCCTCGCAAAAGGCGAGATCGGATCTGCTCAAGCGGCACTACTA 792  
|||||

Db 747 AAATATGAAAAAATCTTGCATGATGCGCAATTTGGATAGTACTATACCAAGATGTT 806  
|||||

QY 793 TGAAGCCCTGGAGACAGCAATGAGCTCGGACAAAGAGCTGTGAGAGAGATGAATT 852  
|||||

Db 807 CGAGCTTATCATTTCCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 866  
|||||

QY 853 TCACATTTGCTGTTTGGCCCAATCCCATCTTCTTGAAGAGAGATCTATCCAGAGATGA 912  
|||||

Db 867 TGAATTTGATGATGATCTTGTATCCAGTATTC--ACGAGATTTATCAAGATTTGAAA 923  
|||||

QY 913 GAAGCAGCTGGGAGAGAGGCTTCAGGCTCAGCTCCGCGGACTTGCATCTCAATGTC 972  
|||||

Db 924 AAGATGAGCGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983  
|||||

QY 973 CGAGAGACCGACTTTCACGCGATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1032  
|||||

Db 984 TTCA---TCAGATTTTGGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1040  
|||||

QY 1033 CGGTCCCGTCCCGAGAGCGGATATCTCGGCGGCTCATGA-----GCACAGGA 1083  
|||||

Db 1041 TCACATGAGCCCTGAAACCTGTTTCAAAACCTGATTCACACATGTGAAATGAAACTGAC 1100  
|||||

QY 1084 GAATTAAGAGAGGAGAGCGGCTGGGCGAGAGAGGCGCTGCGTGGCTGCTGCTGCTG 1143  
|||||

Db 1101 TAATACAGTGTGATCATCTATCGAAGCTGGGAGAGAGAGGAGGCTTTCTTAATTTTACACC 1160  
|||||

QY 1144 GGAATGTTCCGGAAGATCTCGCCGCGGTGTATCGGCTGTAC---GGCAAGCCCATCTA 1200  
|||||

Db 1161 GGAAGCTTACGAAAGTTCTTAACTATTAATTAAGAGAGATTAATTAATTAATTAATTA 1220  
|||||

QY 1201 CATCAACGAGAGAGATGCGGCGGCTGCGAGAGAGAGAGATGATGCGAGAGAGCGCT 1260  
|||||

Db 1221 CATCAAGAAAGAAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1280  
|||||

QY 1261 CAAAGACCCCTCGCATCGGATCTTGAATGATGATGATGATGATGATGATGATGATGAT 1320  
|||||

Db 1281 GAGGAGACATTTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1340  
|||||

QY 1321 TAACAGAGAGCGGCTGCTGATCAAGGGGTAATTTGGTGGGCTGTGCTGATTAATTTGA 1380  
|||||



Db 1341 AGTGAAGATGGTGTGACGTAAAGAGATATTACGATGTGATGATGACAAATTTTGA 1400  
QY 1381 ATGTCAGATGGCTACGAGACCCAGATTTCGG 1410  
Db 1401 ATGGAGCATGTGATACACTGCAAGATTTCG 1430

RESULT 9  
ADA67833  
ID ADA67833 standard; DNA; 1734 BP.  
XX ADA67833;  
AC 20-NOV-2003 (first entry)  
DT 20-NOV-2003 (first entry)  
XX Arabidopsis thaliana gene, SEQ ID 71.  
XX Arabidopsis thaliana gene, SEQ ID 71.  
XX Plant; bacterial infection; fungal infection; viral infection; ds.  
XX Arabidopsis thaliana.  
XX WO2003000898-A1.  
XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI: 2003-175290/17.  
DR Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 6; SEQ ID NO 71; 899bp; English.  
XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.  
XX Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 9.2%; Score 182.8; DB 7; Length 1734;  
Best Local Similarity 49.2%; Pred. No. 8.1e-35;  
Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

QY 79 TCTGCCAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTACGATGGAAGGCGCGT 138  
Db 90 TTTTCTGATGATTTTCATCTTTGGGACAGCTGCTCGGGTTTCAGTACGAGGTGCAAC 149  
QY 139 CAAGAAGGTGGCCCGGCGCGCTCCATCTGGGACACGTACTCCACCTTGGAGCCATCGG 198  
Db 150 AAGTAAGGTGGCAAGTCTCCCACTATAATGAGATCACTTCACGCTATCCAGAAAG 209  
QY 199 CACCAAGCGCGCGGATGGGCTTGGCATCACTACACCGCTACGATGAGAGACTT 258  
Db 210 GACCAAAATGATTAATGAGATGACATTAATTTTATCATCTTTACAAAGATGACAT 269  
QY 259 TGATCTTTGACCAAGTACGGCGCAAAAGGCTTACCGCTTCTCTGTGCTGGTGGGAT 318

Db 270 AAAATTGATGAAGAGCTAAACATGACGCTTTCCGATTTTCAATCTCGGTGCAAAATT 329  
QY 319 CATTCCTCCCTGGCGGACGAGCTGATCCCTTCACACGAGAGGAAATTGATTTTACAGCAA 378  
Db 330 AATACCAATGAAAGCTTAAAGATGAGTAAACAAAGAGGTGACATTTCTACAAAGGA 389  
QY 379 ACTGATGACGCGCTTGTGAGGCGGGATACGCGCTTGGGTGATCTTTGTATCAACACTGGGA 438  
Db 390 TCTCATAGAGAACTTTGCTATATGACATACAACTTCGATGACGCTTATCATTTGGGA 449  
QY 439 TCTGCTCAGGCGCTTACGATGCTATGAGGCTGCTCAAGCTGGAAGAGTCCAGCT 498  
Db 450 CCAACCAAAATCTTTGAGAGACGAAATGTGGCTTTCTAAGCCCTTAAATCTGAGA- 506  
QY 499 GCACTTGAAGCGGATGCGAGGTGTGCTTGAACGTTTGGGACCGGACGCACTG 558  
Db 507 AGATTTTGAAGATTTTGAAGAAATTTGTTGAAGATTTGAGATTAAGTTAAGATGTG 566  
QY 559 GATCACCATCAACGACCCCTGATTCAGGCACTTATGATATGCAACCGGACAAAGC 618  
Db 567 GACAAACATCAACGACCTTATATATGATGCTGTGGGGTTATGATCAAGTAAACAAGC 626  
QY 619 CCGGGCAAGAGC-----AGCATTAACAAGCACTCCACGAGGGCAACACTGCCACTGA 672  
Db 627 GCGTGAACGATGCTCAAAATGGGTAAACGAAAGTGTCAAGCTTGAGATTCGACTACGA 686  
QY 673 GCGGTGGCTGCTGGGAAAGGCCAGATCATGAGCATGCCGCGCGCTGCTACAG 732  
Db 687 GCTTACATTTTTCATCATACACTCTTTTGGCCATGCGCTGCAATGGAAGAAATTTTCG 746  
QY 733 CAGGACCTTTGCGCCCTCGCAAAAGGCGCAAGATCGGATCTCGTCAACGGCGACTACTA 792  
Db 747 AAATGTGAAAAAATCTTCGATGATGCGCAATTTGGATAGTACTATCAACCAAGATGTT 806  
QY 793 TGAAGCTTGGGACGCAATGAGCTTCGGGACAGAGAGCTCTAGAGCGGATGGAATT 852  
Db 807 CGACCTTATATCTTCATTCATCACTGACGATTAAGAGAGCACTTACGAGCTTTCCTT 866  
QY 853 TCACATTTGGCTGTTTGGCAATCCCATCTTTGTAAGAGGACTATCCAGAGACATGAA 912  
Db 867 TGAATTTGATGATGATCTTGAATTCAGTATTC-----ACGAGATTTATCCAGATTTGTA 923  
QY 913 GAGCAGCTGGGCGAGAGGCTTCCAGCCCTCATCTCCGCGACTTTGGCATCTCAATGC 972  
Db 924 AAGTACGCGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983  
QY 973 CGGAGAGCGCACTTTACAGGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1032  
Db 984 TTCA---TCAGATTTTGTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 1040  
QY 1033 CGGTCCGCTCCCGAGAGCGACTATCTCGGCGCATCATGA-----GCACCAAGA 1083  
Db 1041 TCACATTAAGACCTTGAATAAATCTGTTTCAAACTGACCAACATGTAATGAATGAATGAC 1100  
QY 1084 GAATTAAGACCGCAAGCCCGCTTGGCGAGAGAGCGGCTGCTGCGCTGCTGCTGCTG 1143  
Db 1101 TAATCAAGTGTGATCATCATGACCTGCGGAGAAAGAGGCTTCTTAATTTTCAACACC 1160  
QY 1144 GGAATGTTCCGGAAGCATCTCGCCCGGCTGTAAGGCTGTAC---GGCAAGCCCATGTA 1200  
Db 1161 GGAAGGCTTACGAAAGTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1220  
QY 1201 CATCAAGCAAGAGATCCCGCTGCGCTGAGAGAGAACTGACGTGCGAGAGGCGCT 1260  
Db 1221 CATCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1280  
QY 1261 CAACGACCCCTTCGCAATCCGCTTCTTGAATGATGATGATGATGATGATGATGATGAT 1320  
Db 1281 GAAAGCAACATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1340  
QY 1321 TAACCAAGAGCGGCTGCTGCAAGGGTACTTTGCGTGGCGCTTGCATTAATCTTGA 1380



Db 1341 AGTGAAGATGGGTGATGACGTAAAGAGATATTACGATGTCATTGATGACAAATTTGA 1400  
Qy 1381 ATGGTCAGATGGCTACGAGCCAGATTTCGG 1410  
Db 1401 ATGGAGCATGATACACTGCAAGATTTCG 1430

RESULT 10  
ADA70381  
ID ADA70381 standard; DNA: 1533 BP.  
XX  
XX ADA70381;  
AC  
XX 20-NOV-2003 (first entry)  
XX  
XX Rice gene, SEQ ID 3704.  
DE  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX  
XX Oryza sativa.  
OS  
XX MO2003000898-A1.  
PN  
XX 03-JAN-2003.  
PD  
XX 22-JUN-2001; 2001MO-IB001105.  
PF  
XX 22-JUN-2001; 2001MO-IB001105.  
PR  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katsagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;  
XX  
XX WPI; 2003-175290/17.  
DR  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 6; SEQ ID NO 3704; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
XX Sequence 1533 BP; 414 A; 365 C; 402 G; 352 T; 0 U; 0 Other;

Query Match 9.0%; Score 178.6; DB 7; Length 1533;  
Best Local Similarity 49.3%; Pred. No. 8.5e-34;  
Matches 705; Conservative 0; Mismatches 694; Indels 30; Gaps 8;

Qy 73 GCTAGCTCTGCCCAAGACCTTTGAATGGGCTTGCAACGGCGCCCTACCAAGTGAAG 132  
Db 126 GCGAGCTTCCCGAGAGGGGTTCTGTTCGACACGGCGCGCTTACCAAGTGAAGG 185  
Qy 133 CGCGCTCAAGAGAGTGGCGCGCGCGCTTCATCTGGAGACAAGTACCACTTGAGCC 192  
Db 186 CATGGCAAGAGAGGGTGGCGCGCGCGCTTACATCTGGAGAGAGTGAAGAAACAGG 245  
Qy 193 ATGCGGCAACCAAGCGCGCGCGCGCTTGGATGATCACTACCAAGCTTACATGA 252  
Db 246 GAGCATCTCAATTAATATCCACAGCGGATGTGACGGTTGATGATATAGTACAAAGA 305

Qy 253 GGACTTTGATCTCTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTGTGTGTC 312  
Db 306 AGATGTGAACATATATGAAGAACATGGGCTTTGATGGCTATATGATTTTCATCTTGTGTC 365  
Qy 313 GCGGATCATTTCCCTCGGCGGCGAGCTGATCCCGTCAACAGAGAGGAATTGATTTTA 372  
Db 366 AAGATTTTCCCAATGGGACTGGGATG-----GTGAACAGAGAGGAGTTATTTTA 419  
Qy 373 CAGCAACTGATTGACGCGCTGTTGAGCGCGGGATACAGCCTTGGGTGACTTTTGATCA 432  
Db 420 CAACAGTTTAATTAATTAATGATGATGTTAAGAAAGGATCAAAACCTTACCAACTTACCA 479  
Qy 433 CTGGATCTGCTCAGCGCTTACGATCGCTATGAGAGCTGCTCAACCTGAAGAGT 492  
Db 480 CTATGACCTTACATTTGACACTTCATGAGAGAGTACTTAAAGCTTGCTAAGC---CCAACAT 536  
Qy 493 CCAGCTGACTTTGAGCGGTATGCGAGTTGTGCTTTGAACGTTTGGGACCGAGTCCA 552  
Db 537 CGTAGAGCGCTTTGACAGACTACAGATTTCTGCTTCCAGACGTTGAGAGACGGGTGAA 596  
Qy 553 GAACTGGATCAGATCAAGAGACCTGGATTCAGGCGCATCTATGATATGCAACCGGACG 612  
Db 597 GAACTGTTTAACTTCAATGAGCGGAGATGCGTTGCTGCTTGAATATGACATGACTT 656  
Qy 613 CAACGCGCCCGGAGAGAGAGCATTAACAGACCTCCACGAGGCAACCTGCCACTGA 672  
Db 657 CAATGCGCCGGAAGGTCCTGAGCTGTATG-----CAGAGGCAACCTCCAGACAGA 710  
Qy 673 GCCGTGCTTCTGGAAGAGCCCAATCATGAGCATGCGCGCGCTGCGCTTACAG 732  
Db 711 ACCATACCTTGTGCGCCACCATCTCATCTTCTCATGCTGCGCGCTCAAGGATACCG 770  
Qy 733 CAGGACTTTGGCCCGCCGCAAAAGGCGAGATCGGATCTCGCTCAAGCGGACTCTA 792  
Db 771 GAAAGATATCACTTATATCAAAAGGAGATGGAATTCCTTGGATTTTGTCTGTA 830  
Qy 793 TGAGCCCTGGAGACAGCATAGGCTCGGAGCAAGAGCTGTAGCGAGCATGTAAT 852  
Db 831 TGAACATTTAGTACGACGAAATGCT---GATAGGCTGCGAGCAAGAGGCAAGAGATT 887  
Qy 853 TCACTATGCTGTGTTGCCAATCCATCTTCTTGAAGAGAGACTATCAAGAGCATGA 912  
Db 888 TCACTCTGATGATGCTTCTTGAACCCCATTAATC---CATGTGCGTACCATATCTGATGCT 944  
Qy 913 GAGGAGCTGGGAGAGAGGTTTCCAGCCCTCACTCCCGGCACTTTGCCATCTCATATG 972  
Db 945 CGAATTCGTAAAGACAGATGCTCAACTTCACTGATGAGAGATCCAGATGTGAAGA 1004  
Qy 973 CGGAGAGACCGACTTCTACCGCATGTAATTTACTACATCCCACTGCGCGCACTAGA 1032  
Db 1005 CTCAATGATTAATGTTGGCATCAACCACTACCTCTTTCTATCATGAAGAACCTTGCGC 1064  
Qy 1033 CGGTCCGTCGCCGAGACGGAATATCTCGGCGCATCATGAGCAACGAGAAATTAAG- 1091  
Db 1065 ATGGAACCTGACACCAACCACTTATCAGATGATGACTGGCATGTTGGTTTCTATGAACG 1124  
Qy 1092 --ACGGAGCCCGCTTGGCGAGAGAGCGGCTTCGCTGCGCTCGCTCCCGGACAT 1149  
Db 1125 AAACGGGTCGCCATTTGAGACTCAAGCAACTCTTCTGCTTTTACATTTGTGCAATGGG 1184  
Qy 1150 GTTCGGAAGCATCTCGCCCGGCTGTACGCGCTGTACGCGCAAGCCCATCTACATCA 1209  
Db 1185 AATCAACAAAGCTGTGACCTATATGTAAGAAACATATGAAACCTTACATGATCTTTTC 1244  
Qy 1210 GAAACGATGCGCTGCTGCGAGAGAAACATGACGTGCGAGAGGCGCTCAACGACCC 1269  
Db 1245 TGAATAATGATGAGAC---AACCTGGCAACGTACATATCACTCAGGCTGTGATGATAC 1301  
Qy 1270 CTTCGATCGGTAATCTTTGACTCGCACTTGGACTGATTTTCAAGGCCATTTACCA 1329  
Db 1302 AGTAAGATCAAGTACT---ACAGAACTACATCATCACTGAGCTCAAGAGGCGATAGACGA 1358  
Qy 1330 CGCGTGTGTCGTAAGAGGATCTTTGCTGAGGCGTTGCTGATTAATCTGGAATGTACGA 1389



QY 493 CCACTGACCTTTAGAGCGGATATGAGGTTGTCTTGAACGTTTGGGAGCCGATGCA 552  
 Db 657 CAA---CTACTTCAGAGACTAAGCCAACTGTGCTTTGAGCCCTTTGGGAGCCGTGGAA 713  
 QY 553 GAACTGATACCATCAACGAMCCCTGGATTACAGCCATCTATGATATATCCACGGCAG 612  
 Db 714 GCACGTGATACCATCAAGTATCTCGGGCAATGGCAGAAAAGGCTATGAGACGGGCCA 773  
 QY 613 CAACGCCCCGGGAGGAGCAGCATTTAACAGCACTCCACGAGGGCAACCTGCCACTGA 672  
 Db 774 CATTGGCCGGGAGGCTGAAGCTCCGC-----GGCACGGG 806  
 QY 673 GCCGTGCTGCTGTAAGAGCCAGATCATAGCCATGCCGCGCTGGCCGTCTACAG 732  
 Db 807 CCTGTACAGGAGCAGACACCAATCATTTAGGCCCAAGCCAAACCTGGCATTTCTATAA 866  
 QY 733 CAGGACCTTTCGCCCCCTCGCAAAAGGCCAGATGGCATTCGCTCAACGGCGACTACTA 792  
 Db 867 CACCACGTGGCGCAGCAAGGAGGAGGCTGTGGGAATTTCACTGAATGTGACTGGGG 926  
 QY 793 TGAAGCCCTGGGAGAGCAATGAGCTGGGAGCAAGAGGCTGTGAGGAGCATGGAATT 852  
 Db 927 GGAACCTGTGACATTTACTAATCCCAAGACCTAGAGGCTGCCAGAGATACCTACAGTT 966  
 QY 853 TCACATTGGCTGTTTGCCTCAATCCCATCT 881  
 Db 987 CTGCTGGGCTGTTTGCCTCAATCCCATCT 1015

RESULT 12  
 ABL88245  
 ID ABL88245 standard; cDNA; 2016 BP.  
 XX  
 AC ABL88245:

XX 16-MAY-2002 (first entry)  
 DT  
 XX  
 DE Human PRO9820 cDNA sequence SEQ ID NO:147.

XX Human; angiogenesis; cardiant; cytostatic; antiangiogenic; hypotensive;  
 KW vulnary; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;  
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;  
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;  
 KW age-related macular degeneration; arterial restenosis; angina;  
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;  
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;  
 KW wound healing; chromosome mapping; gene mapping; gene; ss.

XX Homo sapiens.

XX MO200200690-A2.

XX 03-JAN-2002.

XX 20-JUN-2001; 2001WO-US019692.

XX 23-JUN-2000; 2000US-0213637P.  
 PR 20-JUL-2000; 2000US-0219556P.  
 PR 25-JUL-2000; 2000US-0220624P.  
 PR 28-JUL-2000; 2000US-0220664P.  
 PR 02-AUG-2000; 2000US-0222655P.  
 PR 17-AUG-2000; 2000US-00643657.  
 PR 23-AUG-2000; 2000WO-US023522.  
 PR 24-AUG-2000; 2000WO-US023328.  
 PR 07-SEP-2000; 2000US-023078P.  
 PR 18-SEP-2000; 2000US-00664610.  
 PR 24-OCT-2000; 2000US-0242922P.  
 PR 08-NOV-2000; 2000US-00709238.  
 PR 08-NOV-2000; 2000WO-US030873.  
 PR 10-NOV-2000; 2000WO-US030873.  
 PR 01-DEC-2000; 2000WO-US032678.

PR 20-DEC-2000; 2000US-00747259.  
 PR 20-DEC-2000; 2000WO-US034956.  
 PR 22-JAN-2001; 2001US-00767609.  
 PR 28-FEB-2001; 2001US-00796498.  
 PR 28-FEB-2001; 2001WO-US005520.  
 PR 01-MAR-2001; 2001WO-US006666.  
 PR 09-MAR-2001; 2001US-00802706.  
 PR 14-MAR-2001; 2001US-00808689.  
 PR 22-MAR-2001; 2001US-00815744.  
 PR 05-APR-2001; 2001US-00828366.  
 PR 10-MAY-2001; 2001US-00854208.  
 PR 10-MAY-2001; 2001US-00854280.  
 PR 25-MAY-2001; 2001US-0086028.  
 PR 25-MAY-2001; 2001US-0086034.  
 PR 25-MAY-2001; 2001WO-US017092.  
 PR 30-MAY-2001; 2001US-00870574.  
 PR 30-MAY-2001; 2001WO-US017443.  
 PR 01-JUN-2001; 2001WO-US017800.

(GENTH ) GENENTECH INC.

PI Baker KP, Ferrara N, Gerber H, Gertlisen ME, Goddard A,  
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,  
 PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

DR WPI; 2002-090516/12.  
 DR P-PSDB; ABB84990.

PT One hundred and eighty seven nucleic acid encoding PRO polypeptides,  
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial  
 PT infarction), endothelial or angiogenic disorders in a mammal.

XX Claim 2; Fig 347; 565pp; English.

PS ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to  
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytostatic,  
 CC antiangiogenic, hypotensive, vulnary and antiarteriosclerotic  
 CC activities, and can be used in gene therapy. The PRO polynucleotides,  
 CC proteins, agonists and antagonists are useful for treating or diagnosing  
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal, e.g.  
 CC cardiac hypertrophy, trauma, cancer, age-related macular degeneration,  
 CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,  
 CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour  
 CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound  
 CC healing. The PRO polynucleotides have applications in molecular biology,  
 CC including use as hybridisation probes, and in chromosome and gene  
 CC mapping. ABL88259 to ABL88267 represent primers and probes used in the  
 CC exemplification of the present invention

XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

XX Query Match 9.0%; Score 178.2; DB 6; Length 2016;

XX Best Local Similarity 54.5%; Pred. No. 1.2e-33;

XX Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

QY 76 AGCTTGCCCAAGCATTTGAATGGGCTTGCAGAGCGCGCTTACAGATGAGGCGC 135  
 Db 237 AATCTTCCCTCTTGGCTTCTCTGGGAGCTGGGCACTTCTGCTTACAGAGGAGCGC 236  
 QY 136 CGTCAAGAGAGTGGCGCGCGCCGTCATCTGGGACAGCTACTGCACTGGAGCCATC 195  
 Db 297 CTGGAGACAGAGAGGAGGAGGCTTAGCATCTGGGACGCTTCAACACAGTGGAGAGG 356  
 QY 196 GCGCACCAACGG--CGCCACGCGCATGTGGCTTGCATCTACCAACGCGTACGATGA 252  
 Db 357 GAAAGTCTTGGGAATGAGCGGCAAGATGAGCTGTGACGCTACTACAAAGTCCAGGA 416  
 QY 253 GACATTTGATCTCTTGAACCAAGTACGCGCAAGGCGCTTCCCTGCTGTGCTGTC 312  
 Db 417 GGCATCATTTCTCTAGAGGAGTGCAGTCAACCACTTACCATTTCTCTGCTTGGCC 476  
 QY 313 GCGGATCATTTCCCTCGCGGCGAGGCTGATCCCGTCAACGAGAGGGAATGATTTTA 372

```

Db      477  CCGGCTCTGCCCCAGCAGCATCCGAGCCGAGCAGGTGAACAAGAGGATTCATTTCTA 536
Qy      373  CAGCAAACTGATGATGACGCCCTCTTGAAGGCGGGGTATCACGCCCTTGGGTGACTTTGATCCA 432
Db      537  CAGTATCTTATGATGATGCGCTTCTAGAGCAGCAACTCACTCCATCTGTGACCTTGACACA 596
Qy      433  CTGGATCTGCTCCTCAGGCGCTTCAAGATCGCTATGAGAGGCTGGCTCAACGTGGAAGAGGT 492
Db      597  CTGGATCTGCGACAGCTGCTCTCAGGTCAAAATACGCTGGGTGGCAGAAATGTGAGCATGGC 656
Qy      493  CCAGCTGGAATTGAGCCGATGCGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCA 552
Db      657  CAA---CTACTTCAGAGACTAGCCCAACCTGTGCTTTAGAGCCTTTGGGAGCCGTTGAA 713
Qy      553  GAACCTGATTCACCATCAAGAACCCCTGATTCAGGCCATTCATGATATGCCACCGGAG 612
Db      714  GCATCTGATTCACGTTTCACTGATCTCGGGCAATGCGCAAAAAGGCTATGAGACGGGCA 773
Qy      613  CAACGCCCCGGGAGCAGCAGCATTTAACAAGCAGCTCCACGAGGGCAACACTGCCACTGA 672
Db      774  CCATGCGCGGGGCTCGAAGCTCCGC-----GGCAGCGG 806
Qy      673  GCCCGGCTCGCTGGAAGGCGCCAGATATAGCCATGCCCGCGCTGGCTTACAG 732
Db      807  CCTGTACAAGGACAGCACAATCATTTAAGGCCACCAAACTGGGCAATTCCTTATTA 866
Qy      733  CAGGGACTTTTGGCCCCCTCGCAAAAGGGCCAGATGGGCACTTCGCTCAACGGCGCACTACTA 792
Db      867  CACCACGTCGGCGCAGCAAGCAAGAGTCTGTGGAAATTTCACTGAACTGTGACTGGGG 926
Qy      793  TGAAGCCTTGGGACAGCAATGAGCCTCGGAGCAAGAGGCTGCTGAGCGACGATGGAATT 852
Db      927  GGAACCTGTGTGACATTAGTAACTCCCAAGAGCCTTAGAGCTGCCAGAGATACCTACAGTT 986
Qy      853  TCACATTCGCTGCTTTGGCAATCCCATCT 881
Db      987  CTGTCTGGGCTGTGTTCACCAACCATT 1015

RESULT 13
ABL95734
ID      ABL95734 standard; cDNA; 2016 BP.
XX      AC      ABL95734;
XX      DT      19-JUL-2002 (first entry)
XX      DE      Human angiogenesis related cDNA PRO9820 SEQ ID NO: 347.
XX      KW      Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
XX      KW      atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
XX      KW      antiarteriosclerotic; gene; ss.
XX      OS      Homo sapiens.
XX      PN      WO200208284-A2.
XX      PD      31-JAN-2002.
XX      PF      09-JUL-2001; 2001WO-US021735.
XX      PR      20-JUL-2000; 2000US-0219556P.
XX      PR      25-JUL-2000; 2000US-0220664P.
XX      PR      28-JUL-2000; 2000WO-US020710.
XX      PR      02-AUG-2000; 2000US-0222695P.
XX      PR      17-AUG-2000; 2000US-00643657.
XX      PR      23-AUG-2000; 2000WO-US023328.
XX      PR      24-AUG-2000; 2000WO-US023322.
XX      PR      07-SEP-2000; 2000US-0230978P.
XX      PR      18-SEP-2000; 2000US-00664610.
XX      PR      18-SEP-2000; 2000US-00665350.

```

```

PR      24-OCT-2000; 2000US-0242922P.
PR      08-NOV-2000; 2000US-00709238.
PR      08-NOV-2000; 2000WO-US030952.
PR      10-NOV-2000; 2000WO-US030873.
PR      01-DEC-2000; 2000WO-US032678.
PR      20-DEC-2000; 2000US-00747259.
PR      20-DEC-2000; 2000WO-US034956.
PR      22-JAN-2001; 2001US-00767609.
PR      28-FEB-2001; 2001US-00796498.
PR      01-MAR-2001; 2001WO-US006520.
PR      09-MAR-2001; 2001US-00802706.
PR      14-MAR-2001; 2001US-00806889.
PR      22-MAR-2001; 2001US-00816744.
PR      05-APR-2001; 2001US-00828366.
PR      10-MAY-2001; 2001US-00854208.
PR      10-MAY-2001; 2001US-00854280.
PR      25-MAY-2001; 2001US-00866028.
PR      25-MAY-2001; 2001US-00866034.
PR      25-MAY-2001; 2001WO-US017092.
PR      30-MAY-2001; 2001US-00870574.
PR      30-MAY-2001; 2001WO-US017443.
PR      01-JUN-2001; 2001WO-US017800.
PR      20-JUN-2001; 2001WO-US019692.

```

```

XX      (GENTH ) GENTECH INC.
PA      (BAKE/) BAKER K P.
PA      (FERR/) FERRARA N.
PA      (GERB/) GERBER H.
PA      (GERR/) GERRIESEN M E.
PA      (GODD/) GODDARD A.
PA      (GODO/) GODOWSKI P J.
PA      (GURN/) GURNEY A L.
PA      (HILL/) HILLAN K J.
PA      (MARS/) MARSTERS S A.
PA      (PANJ/) PAN J.
PA      (PAON/) PAONI N F.
PA      (STEP/) STEPHAN J F.
PA      (WATA/) WATANABE C K.
PA      (WILL/) WILLIAMS P M.
PA      (WOOD/) WOOD W I.

```

```

XX      Baker KP, Ferrara N, Gerber H, Gerriksen ME, Goddard A,
XX      Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,
XX      Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX      WPI; 2002-171999/22.
XX      P-PSDB; ABB95596.

```

```

PT      One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT      useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT      infarction), endothelial or angiogenic disorders in a mammal.

```

```

XX      Claim 1; Fig 347; 567pp; English.

```

```

XX      The present invention provides the protein and coding sequences of human
XX      PRO proteins. These are useful for treating or diagnosing a
XX      cardiovascular, endothelial or angiogenic disorder, including cardiac
XX      hypertrophy, trauma, cancer, age-related macular degeneration,
XX      atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
XX      angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour
XX      angiogenesis (such as breast carcinoma and liver carcinoma) and wound
XX      healing. The present sequence is a coding sequence of the invention

```

```

SQ      Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

```

```

Query Match      9.0%; Score 178.2; DB 6; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.2e-33;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```

```

Qy      76  AGCTCTGCCCAACGACTTGAATGGGCTTGGCAAGCGCCGCTTACCAAGTGAAGGCGC 135
Db      237  AACCTTCCCTTGTGGCTTCTCTGGGGCGGTGGGCAATTCTGCTTACCAAGAGGCGCGC 296

```

QY 136 CGTCAAGAGGTGGCCGCGCCGCTCCATCTGGGACACGACTGCCACCTGAGCCATC 195  
DB 297 CTGGGACACGAGACGGGAAAGGGCTTACGATCTGGGAGCTTTCACACACAGTGGGAAAGG 356  
QY 196 GCGGACCAACGG---CGCCAAACGGCGATGTGGCTTGGGATCACTACACCGCATGATGA 252  
DB 357 GAAAGTCTTGGGAAATGAGACGGGAGATGTAGCTGTGACGGGTACTACAAAGGTCGAGGA 416  
QY 253 GGACTTGTATCTCTTACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTGTGTGTC 312  
DB 417 GGACATCATCTCTGTGGGGAATGACGCTCACGTCMACCATCTACGATTCCTCTGTCTGGCC 476  
QY 313 GCGGATCATTCCTCCGCGCGGACAGGCTGATCCCGTCAACGAGAGGGAATTGATTTTA 372  
DB 477 CCGGCTCTGCGCCACAGGCACTCCAGCGGACAGGTAAACAAGAGGAAATCGAATTTCTA 536  
QY 373 CAGCAAACTGATTTGACCGCCCTGTTGAGCGGGGATACCGCTTGGGTGACTTTTATCA 432  
DB 537 CAGTGATCTTATGATGCTCCCTTCTGACAGCAATCATCTCCATCGTACCTTGGACCA 596  
QY 433 CTGGGATCTGCTTACGGCGCTTACGATCGCTATGAGGCTGGCTCAACGTGGAAAGGT 492  
DB 597 CTGGGATCTGCGACAGCTGCTCCAGTCAATACGCTGGGTGGCAGATGTAGCATGCG 656  
QY 493 CCGGCTGACTTGGAGGGGATGCGAGGTTGTGCTTTGAAAGTTTGGGGACCGAGTCCA 552  
DB 657 CAA---CTACTTCAGAGACTACGCGCAACCTGTGCTTTGAGGCTTTGGGACCGTGTGAA 713  
QY 553 GAATGATGATCACTACACGACGACGACCTGATTCAGGCGCATCTATGATATGCAACCGG 612  
DB 714 GCATGATGATCACGTTCACTGATCTCTCGGGCAATGCGAAGAAAGGCTATGAGAGGCGCA 773  
QY 613 CAACGCCCCGGGACGAGGACGATTAACAGCACTCCACGAGGGCAACCTGCCACTGA 672  
DB 774 CCAATGCCCGGGGCTGAAAGCTCCG-----GGCACCGG 806  
QY 673 GCGGTGCTGCTGGAAAGGCCAGATCATGAGCCATGCCGCGCGCGTCCGCTACAG 732  
DB 807 CTTGTACAGAGGACACACACATCATTTAAGCCACGCAAACTGGCATTTTTATA 866  
QY 733 CAGGACTTTTGCCTCTGCAAAAGGCGCAGATCGGATCTGCTCAACGCGCACTACTA 792  
DB 867 CACACGCTGGCGGACGAAAGGACGAAAGTCTGTGGAAATTTCACTGAATGTGACTGGG 926  
QY 793 TGAGCCTGGGACGACATGAGCTCGGACACAGAGGCTGTGAGCGAGATGAATT 852  
DB 927 GGAACCTGTGACATTTAGTAACTCCCAAGGACCTAGAGGCTGCCAGAGATCTACAGTT 986  
QY 853 TCACATTGGCTGTTGGCAATCCCATCT 881  
DB 987 CTGTCTGGGCTGTTGTCCCAACCCCATTT 1015

## RESULT 14

ACD23979 standard; cDNA; 2016 BP.

ACD23979;

26-AUG-2003 (first entry)

Novel human secreted and transmembrane protein PRO9820 cDNA.

Human; secreted and transmembrane protein; PRO; anti-inflammatory;  
antiarteriosclerotic; cardiant; anti-infertility; anti-HIV; cytosolic;  
antidiabetic; gene therapy; tumour necrosis factor (TNF)-alpha release;  
TNF-alpha release; cell proliferation; cell differentiation;  
gene expression modulator; proteoglycan release; cytokine release;  
tumour; inflammatory disease; organ failure; atherosclerosis;  
cardiac injury; infertility; birth defect; premature aging; AIDS;  
acquired immunodeficiency syndrome; cancer; diabetic complication;  
chromosome mapping; gene mapping; pharmaceutical; diagnostic; biosensor;

KM Bioreactor; tissue typing; gene; ss.  
XX Homo sapiens.  
OS US2003032156-A1.  
XX 13-FEB-2003.  
XX 06-MAY-2002; 2002US-00140474.  
PF 31-MAR-1997; 97WO-US005230.  
XX 12-JUN-1998; 98WO-US012456.  
PR 14-JUL-1998; 98WO-US012452.  
PR 28-AUG-1998; 98WO-US017888.  
XX 10-SEP-1998; 98WO-US018824.  
PR 14-SEP-1998; 98WO-US019093.  
PR 14-SEP-1998; 98WO-US019094.  
PR 14-SEP-1998; 98WO-US019177.  
PR 16-SEP-1998; 98WO-US019330.  
PR 17-SEP-1998; 98WO-US019437.  
PR 07-OCT-1998; 98WO-US021141.  
PR 29-OCT-1998; 98WO-US022991.  
PR 29-OCT-1998; 98WO-US022992.  
PR 20-NOV-1998; 98WO-US024855.  
PR 01-DEC-1998; 98WO-US025108.  
PR 05-JAN-1999; 99WO-US000106.  
PR 08-MAR-1999; 99WO-US005028.  
PR 10-MAR-1999; 99WO-US005190.  
PR 20-APR-1999; 99WO-US008615.  
PR 14-MAY-1999; 99WO-US010733.  
PR 02-JUN-1999; 99WO-US012252.  
PR 01-SEP-1999; 99WO-US020111.  
PR 08-SEP-1999; 99WO-US020594.  
PR 13-SEP-1999; 99WO-US020944.  
PR 15-SEP-1999; 99WO-US021090.  
PR 15-SEP-1999; 99WO-US021547.  
PR 05-OCT-1999; 99WO-US023089.  
PR 29-NOV-1999; 99WO-US028214.  
PR 30-NOV-1999; 99WO-US028313.  
PR 30-NOV-1999; 99WO-US028409.  
PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028634.  
PR 02-DEC-1999; 99WO-US028551.  
PR 02-DEC-1999; 99WO-US028564.  
PR 02-DEC-1999; 99WO-US028565.  
PR 16-DEC-1999; 99WO-US030095.  
PR 20-DEC-1999; 99WO-US030911.  
PR 20-DEC-1999; 99WO-US030999.  
PR 22-DEC-1999; 99WO-US030720.  
PR 30-DEC-1999; 99WO-US031243.  
PR 30-DEC-1999; 99WO-US031274.  
PR 05-JAN-2000; 2000WO-US000219.  
PR 06-JAN-2000; 2000WO-US000277.  
PR 11-FEB-2000; 2000WO-US000376.  
PR 11-FEB-2000; 2000WO-US003565.  
PR 18-FEB-2000; 2000WO-US004341.  
PR 18-FEB-2000; 2000WO-US004342.  
PR 22-FEB-2000; 2000WO-US004414.  
PR 24-FEB-2000; 2000WO-US004914.  
PR 01-MAR-2000; 2000WO-US005004.  
PR 02-MAR-2000; 2000WO-US005046.  
PR 02-MAR-2000; 2000WO-US005841.  
PR 10-MAR-2000; 2000WO-US006319.  
PR 15-MAR-2000; 2000WO-US006884.  
PR 20-MAR-2000; 2000WO-US007377.  
PR 21-MAR-2000; 2000WO-US007532.  
PR 30-MAR-2000; 2000WO-US008439.  
PR 17-MAY-2000; 2000WO-US013705.  
PR 22-MAY-2000; 2000WO-US014042.  
PR 30-MAY-2000; 2000WO-US014941.  
PR 02-JUN-2000; 2000WO-US015264.  
PR 28-JUL-2000; 2000WO-US020710.

PR 11-AUG-2000; 2000WO-US022031.  
PR 23-AUG-2000; 2000WO-US023522.  
PR 24-AUG-2000; 2000WO-US023328.  
PR 08-NOV-2000; 2000WO-US030952.  
PR 10-NOV-2000; 2000WO-US030873.  
PR 01-DEC-2000; 2000WO-US032678.  
PR 20-DEC-2000; 2000WO-US074259.  
PR 28-FEB-2001; 2001US-00796498.  
PR 28-FEB-2001; 2001WO-US006520.  
PR 01-MAR-2001; 2001WO-US006666.  
PR 09-MAR-2001; 2001US-00802706.  
PR 14-MAR-2001; 2001US-00808689.  
PR 22-MAR-2001; 2001US-00816744.  
PR 05-APR-2001; 2001US-00828366.  
PR 10-MAY-2001; 2001US-00854208.  
PR 18-MAY-2001; 2001US-00864289.  
PR 25-MAY-2001; 2001US-00866028.  
PR 25-MAY-2001; 2001US-00866034.  
PR 25-MAY-2001; 2001WO-US017092.  
PR 01-JUN-2001; 2001US-00872035.  
PR 05-JUN-2001; 2001WO-US017800.  
PR 14-JUN-2001; 2001US-00874503.  
PR 19-JUN-2001; 2001US-00886342.  
PR 20-JUN-2001; 2001WO-US019692.  
PR 21-JUN-2001; 2001US-00887879.  
PR 22-JUN-2001; 2001WO-US020116.  
PR 29-JUN-2001; 2001WO-US021066.  
PR 09-JUL-2001; 2001WO-US021735.  
PR 18-JUL-2001; 2001US-00908827.  
PR 06-AUG-2001; 2001US-00924419.  
PR 09-AUG-2001; 2001US-00927796.  
PR 16-AUG-2001; 2001US-00931836.  
PR 19-DEC-2001; 2001US-00028072.  
  
(GETH ) GENENTECH INC.  
XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AJ, Sherwood S;  
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zheng Z;  
XX WPI: 2003-341980/32.  
DR P-PSDBJ; ABO17742.  
XX  
XX New secreted and transmembrane PRO nucleic acids, for treating  
PT inflammation, organ failure, atherosclerosis, cardiac injury,  
PT infertility, birth defects, premature aging, acquired immunodeficiency  
PT syndrome (AIDS), or cancer.  
XX  
XX Claim 2; Fig 253; 660pp; English.  
XX  
XX The invention describes an isolated nucleic acid (I) comprising, or which  
CC has 80 % sequence identity to, or the full-length coding sequence of, one  
CC of 275 nucleotide sequences, and which encodes a corresponding  
CC polypeptide selected from 275 amino acid sequences, where all sequences  
CC are given in the specification. The polypeptide encoded by (I) is used to  
CC detect PRO polypeptides, link a bioactive molecule to a cell expressing a  
CC PRO polypeptide, modulate a biological activity of a cell, stimulate the  
CC release of tumour necrosis factor (TNF)-alpha from human blood, modulate  
CC the uptake of glucose or free fatty acid by cells, stimulate or inhibit  
CC the proliferation or differentiation of cells or gene expression,  
CC stimulate the release of proteoglycans, stimulate the release of cytokine  
CC from peripheral blood mononuclear cells, inhibit the binding of A-peptide  
CC to factor VIIA, or detect the presence of tumour in a mammal. The nucleic  
CC acid and polypeptide encoded by it, are useful for treating inflammatory  
CC diseases, organ failure, atherosclerosis, cardiac injury, infertility,  
CC birth defects, premature aging, acquired immunodeficiency syndrome  
CC (AIDS), cancer, or diabetic complications. The nucleic acid is useful as  
CC hybridisation probes, in chromosome and gene mapping, and in generating  
CC antisense RNA or DNA. The polypeptides are useful as pharmaceuticals,  
CC diagnostics, biosensors or bioreactors. Both are useful in tissue typing.

CC This sequence encodes a novel human secreted and transmembrane PRO  
CC polypeptide  
XX  
SQ Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;  
  
Query Match 9.0%; Score 178.2; DB 7; Length 2016;  
Best Local Similarity 54.5%; Pred. No. 1.2e-33;  
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;  
  
QY 76 AGCTCTGCCCAACGACTTGTAAATGGGCTTTCGCAACGCGCCCTACCGATTCGAAGCGC 135  
DB 237 AACCTTCCTTGGCTTCTCTGGGGCGTGGGAGATTCTGCTTACGACGAGGGCGC 296  
QY 136 CGTAAAGAGGTGGCCGCGCCCTTCATCTGGGACAGTACTGCCACTGAGCCATC 195  
DB 297 CTGGGACCAAGAGGGAAGGGCTTAGCATCTGGAGTGGATTTCACACAGTGGGAAGG 356  
QY 196 GCGCACCAACGG--CGSCAACGGCGATGTGGCTTGGATCACTACACCGCTACAGTGA 252  
DB 357 GAAAGTCTTGGGAATGAGACGGCAGATGTAGCTGTGACGGCTTACTACAAAGTCCAGGA 416  
QY 253 GGAATTGATCTCTTGAACCAAGTACGGCGCAAGGCCCTACCGCTTCTCTGTGTGTC 312  
DB 417 GAAATCATTTCTGCTGAGGAATGCACTGCAACCACTACCGATCTCCCTGTCTGGCC 476  
QY 313 GCGGATCATTCCTCTCGCGCGGAGGCTGATTCCTTCACAGAGAGGAATTAAGTTTA 372  
DB 477 CGGCTCTGCCCAAGGCAATCCGAGCCGAGCGAGTGAACAAAGGAATCAATTTCA 536  
QY 373 CAGCAACTGATTTAGACGCCCTGTGTGAGGCGGGGTATCAAGCTTGGGTGACTTGTACCA 432  
DB 537 CAGTATCTTATGAGATGCCCTTGTGAGCAACATCATCTCCATGTATCTTGACCA 596  
QY 433 CTGGGATCTCTCAGGCGCTTACAGATGCTTATGAGGCTGCTCAAGTGAAGAGT 492  
DB 597 CTGGATCTGCCACACACTGCTCCAGGTCAAAATCGGTGGTGGCAGAAATGTGAGCATGCG 656  
QY 493 CCACTGGAATTTAGAGGGGATAGCGAGTGTGCTTTGAACGTTTGGGGACCGAGTCCA 552  
DB 657 CAA---CTACTTACAGAGACTACCGCAACTGTGCTTTGAGGCTTTGGGACCGTGTGA 713  
QY 553 GAACTGATCAACCATCAAGAGCCCTGATTCAGGCGCATCTATGATATGCCACCGCGAG 612  
DB 714 GCACTGATCAAGTATGATCTCTCGGGCAATGCGAAGAAAGCTATGAGACGGGCA 773  
QY 613 CAACGCCCGGGGAGAGAGCATTTAAACAAGCACTCAACGAGGCAACATGCCACTGA 672  
DB 774 CAAATGCGCGGGCTGAAGCTCGC-----GGCACCGG 806  
QY 673 GCCGTGCTCGCTGGAAGAGCCCAAGATCATGAGCCATGCCGCGCTGACGCTACAG 732  
DB 807 CCGTACAGAGGAGCAGACACCAATCATTTAAGGCCCAACGCAAACTGCAATTTTATA 866  
QY 733 CAGGAACTTTCGCCCTTCGCAAAAGGCCCAATCGGCATCTGCTCAACGCGCACTACTA 792  
DB 867 CACCAAGTGGCGGAGCAACCAAGAGTGTGGGAATTTCACTGAACTGTGAGCTGGGG 926  
QY 793 TGAAGCCCTGGGACAGAAAGAGCTCGGCAAGAGAGGTGCTGAGCGGACGATGAATT 852  
DB 927 GGAACCTTGGAGCAATTAAGTAAACCCCAAGGACCTTAGAGCTGCGAGAGTACTTACAGTT 986  
QY 853 TCACATTGGCTGTTTGGCAATCCCATCT 881  
DB 987 CTGTCTGGGCTGTTTGGCAACCCCATTT 1015  
  
RESULT 15  
ACA67120  
ID ACA67120 standard; cDNA; 2016 BP.  
XX ACA67120;  
AC  
XX 23-JUN-2003 (first entry)

XX cDNA encoding human PRO polypeptide #127.  
DE  
XX  
KM Human; PRO polypeptide; secreted and transmembrane protein;  
KM anti-PRO antibody; diagnostic assay; gene expression; diabetes;  
KM bone disorder; cartilage disorder; rheumatoid arthritis; obesity;  
KM sports injury; osteoarthritis; hyper-insulinaemia; hypo-insulinaemia;  
KM hearing loss; coagulation disorder; stroke; heart attack; cardiac;  
KM antidiabetic; anorectic; vulnery; antiarthritis; osteopathic;  
KM antirheumatic; auditory; cerebroprotective; angiogenic; gene; ss.  
OS  
XX Homo sapiens.  
PN US2003004311-A1.  
XX  
PD 02-JAN-2003.  
XX  
PF 19-DEC-2001; 2001US-00028072.  
XX  
XX 18-JUN-1997; 97US-0049911P.  
XX 26-AUG-1997; 97US-0056974P.  
PR 17-SEP-1997; 97US-0059113P.  
PR 17-SEP-1997; 97US-0059115P.  
PR 17-SEP-1997; 97US-0059117P.  
PR 17-SEP-1997; 97US-0059122P.  
PR 17-SEP-1997; 97US-0059184P.  
PR 18-SEP-1997; 97US-0059263P.  
PR 19-SEP-1997; 97US-0059352P.  
PR 19-SEP-1997; 97US-0059588P.  
PR 24-SEP-1997; 97US-0059836P.  
PR 17-OCT-1997; 97US-0062250P.  
PR 17-OCT-1997; 97US-0062285P.  
PR 17-OCT-1997; 97US-0062287P.  
PR 17-OCT-1997; 97US-0063755P.  
PR 24-OCT-1997; 97US-0062814P.  
PR 24-OCT-1997; 97US-0062816P.  
PR 24-OCT-1997; 97US-0063045P.  
PR 24-OCT-1997; 97US-0063082P.  
PR 24-OCT-1997; 97US-0063127P.  
PR 27-OCT-1997; 97US-0063327P.  
PR 27-OCT-1997; 97US-0063329P.  
PR 28-OCT-1997; 97US-0063550P.  
PR 28-OCT-1997; 97US-0063561P.  
PR 29-OCT-1997; 97US-0063704P.  
PR 29-OCT-1997; 97US-0063733P.  
PR 29-OCT-1997; 97US-0063735P.  
PR 29-OCT-1997; 97US-0063738P.  
PR 03-NOV-1997; 97US-0064248P.  
PR 03-NOV-1997; 97US-0064809P.  
PR 12-NOV-1997; 97US-0065186P.  
PR 17-NOV-1997; 97US-0065846P.  
PR 21-NOV-1997; 97US-0066364P.  
PR 24-NOV-1997; 97US-0066453P.  
PR 24-NOV-1997; 97US-0066511P.  
PR 24-NOV-1997; 97US-0066707P.  
PR 11-DEC-1997; 97US-0069212P.  
PR 11-DEC-1997; 97US-0069278P.  
PR 16-DEC-1997; 97US-0069694P.  
PR 23-JAN-1998; 98US-0072320P.  
PR 04-FEB-1998; 98US-0073612P.  
PR 09-FEB-1998; 98US-0074086P.  
PR 09-FEB-1998; 98US-0074092P.  
PR 12-MAR-1998; 98US-0077791P.  
PR 20-MAR-1998; 98US-0078910P.  
PR 27-MAR-1998; 98US-0079294P.  
PR 27-MAR-1998; 98US-0079663P.  
PR 31-MAR-1998; 98US-0079728P.  
PR 12-JUN-1998; 98US-0080165P.  
PR 14-JUL-1998; 98WO-US012456.  
PR 28-AUG-1998; 98WO-US014552.  
PR 10-SEP-1998; 98WO-US017888.  
PR 98WO-US018824.

PR 14-SEP-1998; 98WO-US019093.  
PR 14-SEP-1998; 98WO-US019094.  
PR 14-SEP-1998; 98WO-US019177.  
PR 16-SEP-1998; 98WO-US019330.  
PR 17-SEP-1998; 98WO-US019437.  
PR 07-OCT-1998; 98WO-US021141.  
PR 29-OCT-1998; 98WO-US022991.  
PR 29-OCT-1998; 98WO-US022992.  
PR 20-NOV-1998; 98WO-US024855.  
PR 01-DEC-1998; 98WO-US025108.  
PR 05-JAN-1999; 99WO-US000106.  
PR 08-MAR-1999; 99WO-US005028.  
PR 10-MAR-1999; 99WO-US005190.  
PR 20-APR-1999; 99WO-US008615.  
PR 14-MAY-1999; 99WO-US010733.  
PR 02-JUN-1999; 99WO-US012252.  
PR 01-SEP-1999; 99WO-US020111.  
PR 08-SEP-1999; 99WO-US020594.  
PR 13-SEP-1999; 99WO-US020944.  
PR 15-SEP-1999; 99WO-US021547.  
PR 05-OCT-1999; 99WO-US023089.  
PR 29-NOV-1999; 99WO-US028214.  
PR 30-NOV-1999; 99WO-US028313.  
PR 30-NOV-1999; 99WO-US028409.  
PR 01-DEC-1999; 99WO-US028301.  
PR 01-DEC-1999; 99WO-US028304.  
PR 02-DEC-1999; 99WO-US028551.  
PR 02-DEC-1999; 99WO-US028564.  
PR 02-DEC-1999; 99WO-US028565.  
PR 16-DEC-1999; 99WO-US030095.  
PR 20-DEC-1999; 99WO-US030911.  
PR 20-DEC-1999; 99WO-US030999.  
PR 30-DEC-1999; 99WO-US031243.  
PR 30-DEC-1999; 99WO-US031274.  
PR 05-JAN-2000; 2000WO-US000219.  
PR 06-JAN-2000; 2000WO-US000277.  
PR 11-FEB-2000; 2000WO-US003565.  
PR 18-FEB-2000; 2000WO-US004341.  
PR 18-FEB-2000; 2000WO-US004342.  
PR 22-FEB-2000; 2000WO-US004414.  
PR 24-FEB-2000; 2000WO-US004914.  
PR 24-FEB-2000; 2000WO-US005004.  
PR 01-MAR-2000; 2000WO-US005601.  
PR 02-MAR-2000; 2000WO-US005746.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;  
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WT, Zhang Z;  
XX  
XX WPI; 2003-352836/33.  
DR P-PSDB; ABU80996.  
DR  
XX  
PT New isolated PRO polypeptide useful for treating diabetes, rheumatoid  
PT arthritis, sports injuries, obesity, hearing loss in mammals, stroke, or  
PT heart attack.  
XX  
XX  
PS Claim 2; Fig 253; 643pp; English.  
XX  
XX The present invention relates to the isolation of novel human PRO  
XX polypeptides, and the polynucleotide sequences encoding them. The PRO  
XX polypeptides are secreted and transmembrane proteins. The PRO  
XX polypeptides and polynucleotides are useful for preparing a medicament  
XX useful in the treatment of diabetes, bone and/or cartilage disorders  
XX (e.g. rheumatoid arthritis, sports injuries, osteoarthritis), obesity,  
XX hyper- or hypo-insulinaemia, hearing loss, and coagulation disorders  
XX (e.g. stroke, heart attack). Anti-PRO antibodies are useful in diagnostic  
XX assays for PRO, by detecting its expression in specific cells, tissues or  
XX serum, and for affinity purification of PRO from recombinant cell culture  
XX or natural sources. ACA65994-ACA67268 represent cDNA sequences encoding

the human pro polypeptides of the invention. Note: The sequence data for this patent was obtained in electronic format directly from the USPIO web site at [seqdata.uspio.gov/psipidEntry.html](http://seqdata.uspio.gov/psipidEntry.html)

Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

```
Query March 9.0%; Score 178.2; DB 7; Length 2016;
Best Local Similarity 54.5%; Pred. No. 1.2e-33;
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

QY 76 AGCTTGTCCCAACGACTTTGAATGGGCTTGGCAACGGCCGCTTACCAAGATCGAAGGCG 135
DB 237 AACCTTCCCTCTTGAGCTTCTCTGGGGGTGGGCAAGTTCTGCTTACCAAGAGGCGC 296
QY 136 CGTCAAAAGAGTGGCGCGCGCCGCTCCATCTGGGACACGTACTGCCACCTGAGCCATC 195
DB 297 CTGGGACCAAGACGGGAAAGGAGCTTACATCTGGGACGCTTTCACACACAGTGGAAAGG 356
QY 196 GCGCACCAACG---CGCCAAAGGAGATGGGCTTGGCATCTACCAACCGCTACGATGA 252
DB 357 GAAAGTGCTGGGAATGAGACGGCAGATGTAGCCTGTGACGCTTACCAAGTCCAGGA 416
QY 253 GGACTTTGATCTCTTGACCAAGTACGGGGCAAAAGGCTTACCGCTTCTCTGTGATGTC 312
DB 417 GGACATCATTTCTGTGAGAGGAACTGCACGTCAACCACTACCGATTTCTCTGTGGCC 476
QY 313 GCGGATCATTTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGGAATTGATTTTA 372
DB 477 CCGGCTCTCGCCACAGGATCCGAGCCGAGCAGGTGAACAAAGAGGAATCGAATCTA 536
QY 373 CAGCAAATGATTTAGACGCTTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACA 432
DB 537 CAGTGAATCTTATCGATGCGCTTCTGACGACAAACATCACTCCATCGTACCTTGACCA 596
QY 433 CTGGGATCTGCTCAGGCGCTTCAAGATGCTATGAGGCTGGCTCAAAGTGAAGAGT 492
DB 597 CTGGGATCTGCCACAGCTGCTCAGGTCAATACGTTGGGTGCAAAATGTAGCATGGC 656
QY 493 CCAGCTGACCTTTGAGCGGTATGCGAGGTTGTGCTTTGAACGTTTGGGGACCGAGTCCA 552
DB 657 CAA---CTACTTCAGAGACTACGCCCAACTGTGCTTTGAGGCTTTGGGGACCGTGTGA 713
QY 553 GAATGGATCACCATCAAGAMCCCTGATTCAGGCCATCTATGATATGACACCGGCGAG 612
DB 714 GCACTGATACGTTCAATGATCTCTGGGCAATGSCAGAAAAGGCTATGAGACGGGCCA 773
QY 613 CAAGGCCCGGAGGAGGACGATTAACAAGCACTCCACGAGGGCAACACTGGCACTGA 672
DB 774 CCATGGGCGGGGCTGAAGCTCCG-----GGCACCGG 806
QY 673 GCCGTGGCTGCTGGAAGGCCCAAGATCATGAGCCATGCCCGGCGGTGCGCTACAG 732
DB 807 CCTGTACAAAGCAGACACACATCATTAAGGCCCAAGCCAAACCTGGCATTTCTTATAA 866
QY 733 CAGGGAATTTCGCCCTCGCAAAAGGCGCAGATCGGCATCTCGCTCAACGCGCATTA 792
DB 867 CACCACTGGCGCAGACAGCAAGGTCTGTGGGAATTTCACTGAATGTGACTGGGG 926
QY 793 TGAAGCCTGGAGACGCAATGAGCTCGGACAAAGAGGCTGTGAGCGACGATGGAATT 852
DB 927 GGAACCTGTGGACATTGTAACTCCCAAGACCTTAAAGAGCTGCCAGAGATACCTA 986
QY 853 TCACATGGCTGGTTTGGCAATCCCATCT 881
DB 987 CTGTCTGGGCTGTTTGCACACCCCATTT 1015
```

Search completed: March 28, 2004, 17:03:07  
Job time : 801.658 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:52:22 ; Search time 144.443 Seconds  
(without alignments)  
7649.443 Million cell updates/sec

Title: US-10-026-140-1

Perfect score: 1991  
Sequence: 1 agcagcagtcgcacacagcagc.....ttttaaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	7.5	1521	US-09-914-841A-14	Sequence 14, Appl
2	139.4	7.0	1467	US-09-489-039A-2386	Sequence 2386, Ap
3	130.8	6.6	1317	US-09-134-078-2	Sequence 2, Appl
4	128.2	6.4	3163	US-09-344-510B-6	Sequence 6, Appl
5	128.2	6.4	3435	US-09-344-510B-7	Sequence 7, Appl
6	127.8	6.4	1650	US-09-344-510B-9	Sequence 9, Appl
7	127.8	6.4	5032	US-09-344-510B-8	Sequence 8, Appl
8	126.2	6.3	1287	US-09-914-841A-2	Sequence 2, Appl
9	124.8	6.3	1542	US-09-122-230-6	Sequence 6, Appl
10	124.8	6.3	1909	US-09-122-230-6	Sequence 6, Appl
11	118.6	6.0	1829	US-08-688-988-9	Sequence 9, Appl
12	111.4	5.6	1545	US-08-400-275-17	Sequence 17, Appl
13	104.8	5.3	1458	US-09-134-000C-2890	Sequence 2890, Ap
14	101.4	5.1	1437	US-09-107-532A-1268	Sequence 1268, Ap
15	100	5.0	7218	US-08-232-463-14	Sequence 14, Appl
16	95.6	4.8	1410	US-09-489-039A-1053	Sequence 1053, Ap
17	93.8	4.7	1407	US-08-688-988-7	Sequence 7, Appl
18	90.4	4.5	2435	US-09-306-593-1	Sequence 1, Appl
19	88.6	4.5	1931	US-09-431-470-1	Sequence 1, Appl
20	88.6	4.4	1931	US-09-431-470-3	Sequence 3, Appl
21	87.6	4.4	454	US-09-615-192A-166	Sequence 166, App
22	87.6	4.4	454	US-09-169-789-166	Sequence 166, App
23	85.2	4.3	457	US-08-975-316-79	Sequence 79, Appl
24	85.2	4.3	457	US-09-615-192A-79	Sequence 79, Appl
25	85.2	4.3	457	US-09-169-789-79	Sequence 79, Appl
26	85.2	4.3	470	US-09-615-192A-141	Sequence 141, App
27	85.2	4.3	470	US-09-169-789-141	Sequence 141, App

28	84.8	4.3	3460	US-09-344-510B-10	Sequence 10, Appl
29	84.2	4.2	1404	US-09-489-039A-4718	Sequence 4718, Ap
30	84.2	4.2	1455	US-09-489-039A-4800	Sequence 4800, Ap
31	79.4	4.0	1443	US-09-489-039A-1986	Sequence 1986, Ap
32	74.8	3.8	1635	US-08-961-527-74	Sequence 74, Appl
33	71	3.6	1428	US-09-489-039A-2983	Sequence 2983, Ap
34	69.4	3.5	1479	US-09-134-000C-289	Sequence 289, App
35	68.4	3.4	286	US-09-313-294A-7321	Sequence 7321, Ap
36	68	3.4	3615	US-08-920-812-17	Sequence 17, Appl
37	68	3.4	3615	US-08-920-812-17	Sequence 17, Appl
38	68	3.4	3615	US-08-921-177-17	Sequence 17, Appl
39	68	3.4	3615	US-08-962-577C-17	Sequence 17, Appl
40	68	3.4	3615	US-08-920-828-17	Sequence 17, Appl
41	63	3.2	11303	US-08-961-527-115	Sequence 115, App
42	62.6	3.1	15231	US-09-128-155-16	Sequence 16, Appl
43	62.6	3.1	176373	US-09-128-155-17	Sequence 17, Appl
44	61	3.1	1359	US-09-019-095A-7	Sequence 7, Appl
45	58	2.9	494	US-08-332-766A-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1  
US-09-914-841A-14  
Sequence 14, Application US/09914841A  
Patent No. 6645750  
GENERAL INFORMATION:  
APPLICANT: Amamo Enzyme Inc.  
TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE  
FILE REFERENCE: 066072  
CURRENT APPLICATION NUMBER: US/09/914, 841A  
CURRENT FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: P. Hei. 11-056299  
PRIOR FILING DATE: 1999-03-04  
PRIOR APPLICATION NUMBER: PCT/JP00/01242  
PRIOR FILING DATE: 2000-03-02  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 1521  
TYPE: DNA  
ORGANISM: Camellia var. sinensis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1521)  
OTHER INFORMATION:  
NAME/KEY: mat peptide  
LOCATION: (235)..()  
OTHER INFORMATION:  
US-09-914-841A-14  
Query Match 7.5%; Score 150; DB 4; Length 1521;  
Best Local Similarity 52.2%; Pred. No. 1.5e-28;  
Matches 455; Conservative 1; Mismatches 401; Indels 15; Gaps 5;  
81 TGGCCACAGACTTGGATGGGCTTGGCAAGCCCGCTTACCAAGTGAAGGCGCGTCA 140  
116 TTCTGATGTTTGTGTTTGGAGCTGCTTCTTCCCTTACAGTTGAAGGTGGCCA 175  
141 AAGAAGTGGCCGCGCCGCTTCATCTGGAGACAGTACGCCACCTGAGGCCATCCGCA 200  
176 AGAAGGTGGGAAGGCCCAATTTTGGGATCCTTCACTCAATGATTTCCAGTAA 235  
201 --CAAGCGCGCAAGCGGATGCTTGGCATCACTACCAAGCCCTACGATGAGACT 257  
236 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 295  
258 TTTATCTTTGACCAAGTACGAGCAAGCCCTTCCCTTCTGCTGCTGCTGCTGCT 317  
296 TGAAGGTGCTGAAGTTTATGATGATGATGATGATGATGATGATGATGATGATG 355  
318 TCATTCCCTCGGCGGACGCTGATCCGTCACAGAGGAGGATTTGATTACAGCA 377

Db 356 TATTACCTGGGGGAAAGCTTAGCGAGAGTGAACAAGAGTATCCGCTTCAACACA 415  
Qy 378 AACGTATACGCCCTGTGTAGAGCGGGGATATACAGCTTGGTACTTTGACCACTGGG 437  
Db 416 ATGTCAATCAATACCTTTATGCAAAAGGTATCAACCTTTTATTAACAATCTTCACTGGG 475  
Qy 438 ATTCGCTCAGCGCTTCAAGATCGCTATGAGAGGCTGCTCAACGTGGAAGGCTCCAGC 497  
Db 476 ATCTTCCCAAGCCCTAGAGATGATATGAGAGCTTTTAAAGC---CCACATATTGGA 532  
Qy 498 TGGACTTTAGCGGTATGAGAGTGTGCTTTGAAAGTTTGGGAGCCAGTCCAACT 557  
Db 533 ACGATTTCCGGGATTTTGAAGAGCTGTGCTTCAAGAGATTGTTGACCGAGTTAAACAT 592  
Qy 558 GGATCAACATCAACGAMCCCTGATTCAGAGCCATCTATGATATGCAACCGGACGAACG 617  
Db 593 GGAATCAATGATGATGACCATGCTTACTCTATGAGGGGTTATGATGACAGGCTCTAG 652  
Qy 618 CCCCAGGAGAGCAGCATTTAACAAGCACTC---CACGAGGGCAACAACCTGCACTGAGC 674  
Db 653 CACCGGGCCGTTGTTGGCTTTATGCACTTTTGCCTTAAGGAAATTTGGAGCTGAGC 712  
Qy 675 CGTGGCTGCTGGAAGGCCAGATCATGAGCCATGCCCGCGCTGAGCCCTTACAGCA 734  
Db 713 CTTATATAGTTAACCCACAATTTGCTTCTCATGCTGCTGCTGTAACATATACAGG 772  
Qy 735 GGGACCTTTCGCCCCGCAAAAAGGCGCATGCGCATCTCGTCAACGGCACTACTATG 794  
Db 773 AGAATATACAGCATATCAAAAAGGCGCATGAGATTAACACTGATGACTTATGATGA 832  
Qy 795 AGCCCTGGAGACGAATGAGCCTCGGGAACAAGAGCTGTGAGCGACGATGGAATTC 854  
Db 833 TTCCCT---ACTCCATTCGAAAGCCGACAGGATGACGACAACGCCCTTATTTCA 889  
Qy 855 ACATTGCTGCTTGGCCATCCATCTTCTTGAAGAAGACTATCCAGAGCATGAGA 914  
Db 890 TGTATGATGATGTTATTTAGCCATTAACTT---TGTGATATATCAAAAAGCATGCGCTA 946  
Qy 915 AGCAGCTGGGCGAGAGCTTCAGCCCTCACT 946  
Db 947 GACTGTTGGTAAAGTTTACCAGGTTCACT 978

RESULT 2  
US-09-489-039A-2386  
; Sequence 2386, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; PRIORITY FILING DATE: 2000-01-27  
; PRIORITY FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2386  
; LENGTH: 1467  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2386

Query Match 7.0%; Score 139.4; DB 4; Length 1467;  
Best Local Similarity 56.1%; Pred. No. 7.5e-26;  
Matches 307; Conservative 1; Mismatches 227; Indels 12; Gaps 2;

Qy 77 GCTTCCCAACGAGCTTGAATGGGGCTTGGCAAGCGCCGCTTACAGATGAGAGCGCC 136  
Db 103 GCTTTCCGACAAATTTCTTATGGGGTGGCGCTACCGACGATTCAGGTTGAGGAGGA 162  
Qy 137 GTCAAAAGAGTGGCCCGCGCTCATCTGGGACACGTAAGTCCACCTGAGACCATCG 196

Db 163 CACACGCCAGCGTAAAGACCTGATCTGGATATCTACTCTCATCTGCGGGTACC 222  
Qy 197 GCACCAACCGCGCAACCGGATGATGCTTGGCATCATACACCGCTACGATGAGAC 256  
Db 223 ACTTTGAAGCACACCGGATATTTGCGTGCACACTATACAGCTTTTCGGAGAAC 282  
Qy 257 TTTATCTTTGACCAAGTACGCGGCAAGGCTTCTCTCTTCTGCTGCTGCGG 316  
Db 283 GTGCGTTAATGCGCGATGAGGCTGCAAAAGCATATGATTTCTCATCTCTGCCCCG 342  
Qy 317 ATCATTTCCCTCGGCGGAGCTGATCCGTCACAGAGGAGAAATGATTTTACAGC 376  
Db 343 CTGCTGCC-----CGCCGGCGCGGTAAAGTAAAGAGTCCAGTTTACAGC 396  
Qy 377 AAATGATTAACGCTTGTGAGCGGGATACAGCTTGGTGAATTTGATCACTGG 436  
Db 397 GATCTGATGACGAACGCTGCGCACATATGAGCCGATGATCACCTCTATCACTGG 456  
Qy 437 GATCTGCTGAGCGCTTTCAGATCGCTATGAGAGCTGCTCAACGTGGAAGGTTCCAG 496  
Db 457 GATCTGCGAGGCGCTTCAGG-----ATGAGGGCGGCTGGGAAGCGGCACTACGCC 510  
Qy 497 CTGACTTTGAGCGGTATGCGAGGTTGTGCTTGAAGCTTTGGGACCGAGTCCAGAAC 556  
Db 511 GAAGCCTTGGCGAGTACGCGCGCTGTGCTATGCGCGCTTTGTTACAGGGTGAAGCTG 570  
Qy 557 TGGATCAATCAACGAMCCCTGATTCAGGCCATCTATGATATGCCACCGGACGAC 616  
Db 571 TGGGCTAATCTTCAACGAACCATTTGTTCAATGAGGACAGGCTAATTAACGGCTTCAAT 630  
Qy 617 GCCCGCG 623  
Db 631 CCCCCG 637

RESULT 3  
US-09-134-078-2  
; Sequence 2, Application US/09134078  
; Patent No. 6368844  
; GENERAL INFORMATION:  
; APPLICANT: Bylina, Edward J.  
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
; STREET: 4365 Executive Drive, Suite 1600  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/134,078  
; FILING DATE: 13-AUG-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,026  
; FILING DATE: 10-OCT-1997  
; APPLICATION NUMBER: 60/056,916  
; FILING DATE: 06-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hallie, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/024002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 858/677-1456  
; TELEFAX: 858/677-1465  
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1317 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 FEATURE:  
 NAME/KEY: Coding Sequence  
 LOCATION: 1..1314  
 US-09-134-078-2

Query Match 6.4%; Score 130.8; DB 4; Length 1317;  
 Best Local Similarity 56.4%; Pred. No. 1.1e-23;  
 Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

```

QY 77 GCTCTGCCCCAACGACCTTTGAATGAGGCTTGGCAACGGCCCTTACCGCATCGAAGGGCC 136
DB 16 GATTTTCGAAAAGATTTTATCTTCGAAACGGCTACCGACATACCGAATTGAAGGTGCA 75
QY 137 GTCAGAGAGAGTGGCGCGCGCCGTCATCTGGAGACAGTACGCCACTGGAGCCATCG 196
DB 76 GCAGACGAGAGATGGCAGAGGGCCATCAATTGGAGTGTCTTTCACACAGCCCTGGCAA 135
QY 197 CGCAGCAACGGCGCCCAACGGCGATGTGGCTTGGCATCTACCAACCGCTACGATGAGAC 256
DB 136 ACCCTGAACGGTGACACAGAGACGTTGCGTGACCATTAATCACCGATACAGGAAGAT 195
QY 257 TTGATCTCTTGAACCAAGTACGGCGCAAGAGCCCTTCTCTTGTGTGTGTCGGG 316
DB 196 ATCCAGCTGATGAAGAAAGATAGGATTAAGACGCTTACAGGTTCTCTATCTCTGGCCGA 255
QY 317 ATCATTCCTCCCTCGCGGCGAGCTGATCCGCTCAACAGAGAGGAATTGAGTTTACAGC 376
DB 256 ATTATGCCAGATGG-----GAGAAATCAACCAAAAGGTTGGATTCTTCTACAC 306
QY 377 AAATGATGACGCCCTGTTGAGCGGGATATCAACGCTTGGTGAATTGTAACACTGG 436
DB 307 AGACTCGTTGATGAGCTTTTGAAGAAATGATATCATACATTCGTAACACTCTACCTGG 366
QY 437 GATCTGCTTACGGCGCTTCAACGATCGCTATGAGAGGCTGCTCAACCGTGAAGAGTCCAG 486
DB 367 GACTTACCTTACCCACTTATTAAGAAA--AAGGTGAGTGGCTTAAACC--AGATATAGCG 420
QY 497 CTGACCTTGAAGGATGAGAGGTTGCTTGAAGCTTTTGGGAGCGAGTCCAGAAC 556
DB 421 CTCTATTTCAGACATACGCAACGTTTATGTTCAACGAACCTCGTATCTGTGAACAT 480
QY 557 TGGATCACCATCAACGAMCCCTGATTCAGGCGCATCTATGATATGCCACCGGACGAC 616
DB 481 TGGATTACACTGAACGAAACCATGTGTCTTCTTCTCGGTTATTAACAGGAGAGCAT 540
QY 617 GCGCCGGG 624
DB 541 GCGCCGGG 548

```

# RESULT 4

US-09-144-510B-6  
 Sequence 6, Application US/09344510B  
 Patent No. 6579850  
 GENERAL INFORMATION:  
 APPLICANT: Nabeshima, Youichi  
 Kuroo, Makoto  
 Sekine, Susumu  
 Iida, Akihito  
 TITLE OF INVENTION: No. 6579850el Polypeptide, No. 6579850el DNA and No. 657985  
 NUMBER OF SEQUENCES: 38  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
 STREET: 30 Rockefeller Plaza  
 CITY: New York  
 STATE: New York  
 COUNTRY: United States

ZIP: 10112-3801  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
 COMPUTER: Compaq Deskpro EN  
 OPERATING SYSTEM: Windows 98  
 SOFTWARE: Wordpad  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/344,510B  
 FILING DATE: 25-Jun-1999  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/JP97/04585  
 FILING DATE: 12-DEC-1997  
 APPLICATION NUMBER: JP 347871  
 FILING DATE: 26-DEC-1996  
 APPLICATION NUMBER: JP 205815  
 FILING DATE: 31-JUL-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Perry, Lawrence S.  
 REGISTRATION NUMBER: 31865  
 REFERENCE/DOCKET NUMBER: 766.32  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 218-2100  
 TELEFAX: (212) 218-2200  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3163  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 ORIGINAL SOURCE:  
 ORGANISM: human  
 IMMEDIATE SOURCE:  
 LIBRARY: kidney  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 9..3047  
 IDENTIFICATION METHOD: E  
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
 US-09-144-510B-6

Query Match 6.4%; Score 128.2; DB 4; Length 3163;  
 Best Local Similarity 52.1%; Pred. No. 7.3e-23;  
 Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

```

QY 191 CCATCGGCGCAACAGCGCCCAACGGCGATGTGGCTTGGCATCTACCAACCGCTACGAT 250
DB 348 CCGTCGCGCGTGACGCCCGCACCGGGGAGCTAGCCAGGACAGCTAACAAACGCTTTC 407
QY 251 GAGACCTTGGATCTTGAACCAAGTACGGCGCAAGAGCCCTACCGCTTCTGTGTG 310
DB 408 CGGACACGAGAGCGCTGCGCGAGCTCGGGGTACTCATCACTTCCTTCATCTGTGG 467
QY 408 CGGACACGAGAGCGCTGCGCGAGCTCGGGGTACTCATCACTTCCTTCATCTGTGG 467
DB 311 TCGCGATCATTCCTCGCGCGAGCTGATCCGCTCAACGAGAGGGAATTGAGTTT 370
QY 468 GCGGAGTGTCTCCCAATGAGCAGCGGGGGGTCC--CAACGAGAGGGGCTGCGCTAC 524
DB 371 TACAGCAAACTGATTGAACGCGCTTGGAGCGGGGTATCAACGCTTGGTGAATTGAC 430
QY 525 TACCGGCGCTGTGAGAGCGCTGCGGAGCTGGGAGCTGGGAGCTGGTGCACCTGTAC 584
DB 431 CACTGGATCTGCTCAGAGCGCTTCAAGATCGCTATGAGAGGCTGGCTCAACGTGAAGAG 490
QY 585 CACTGGACCTGCGCCAGCGCTGCGAGAGCTTACGCGGCTGGGCAACCGCGCCCTG 644
DB 491 GTCAGCTGACTTTGAGCGGATGAGAGTGTGCTTTGAAGCTTTTGGGACCGAGTC 550
QY 645 GCGGACC--ACTTACGGGATTAACGGGAGCTGTCTTCGACACTTCGCGGCTGACGTC 701
DB 551 CAGAACTGATCACTCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 610
QY 702 AAGTACTGATCACTCAACCAACCCCTACGTTGTTGGCTGGCAGCGCTTACCGCACCGG 761

```



RESULT 6  
US-09-344-510B-9  
Sequence 9, Application US/09344510B  
Patent No. 6579850  
GENERAL INFORMATION:  
APPLICANT: Nabeshima, Youichi  
Kuroo, Makoto  
Sekine, Susumu  
Iida, Akihito  
TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10112-3801  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
COMPUTER: Compaq Deskpro EN  
OPERATING SYSTEM: Windows 98  
SOFTWARE: Wordpad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/344,510B  
FILING DATE: 25-Jun-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/04585  
FILING DATE: 12-DEC-1997  
APPLICATION NUMBER: JP 347871  
FILING DATE: 26-DEC-1996  
APPLICATION NUMBER: JP 205815  
FILING DATE: 31-JUL-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Perry, Lawrence S.  
REGISTRATION NUMBER: 31865  
REFERENCE/DOCKET NUMBER: 766.32  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 218-2100  
TELEFAX: (212) 218-2200  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1650  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: mouse  
IMMEDIATE SOURCE:  
LIBRARY: kidney  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1650  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-09-344-510B-9  
Query Match 6.4%; Score 127.8; DB 4; Length 1650;  
Best Local Similarity 53.1%; Pred. No. 7e-23;  
Matches 388; Conservative 0; Mismatches 307; Indels 36; Gaps 4;

Db 466 GCGGGGTGCTCCCAATGGACCGCGGCATCTC---CAACGCGAGGGGCTGCGTAC 522  
Qy 371 TACAGCAACTGATATGACGCCCTGTTAGAGCGGGATACAGCCTTGAGTACTTTTAC 430  
Db 523 TACCGGGGTGCTGAGCGCGCTGCGGACCTGGCGCTGACCGGTGATACCTTTAC 582  
Qy 431 CACTGGATCTGCTCAGCGCTTACAGATCGCTATGAGAGCTGAGCTCAACGTGAAGAG 490  
Db 583 CATGGACCTGCGACAGCGCTGAGAGACCTATGCGGATGGCCATCGGCCCTG 642  
Qy 491 GTCCAGCTGACCTTGAACGGTATGCGAGCTTGCTTTGAAGCTTTGGGACCGAGTC 550  
Db 643 GCGGACC---ATTCAAGGATTAATGCGAGCTGCTTCGCGCACTTCGGTGGTCAAGTC 699  
Qy 551 CAGAACTGATACACTCAACGAMCCTGATTCAGGCCATCTATGATATGCCACCGGC 610  
Db 700 AAGTACTGATACACTCAACCCCTACGTGATGCTTGGCAGCGGTATGCGACCGGG 759  
Qy 611 AGCAAGCGCCCGGCGAGAGCAGATTAACAGCACTCCACCGAGGCAACATGCCACT 670  
Db 760 GCGCTGCGCCCGGCGAGGGGAGCTCAGGCTCGG----- 797  
Qy 671 GAGCGTGGCTCGCTGAAAAGGCCAGATCATGAGCCATCCGCGCGCTGCGCTAC 730  
Db 798 -----GTACTGGTGGCCCAACACTTCTTTGGCTCATGCAAAAGCTGGCATCTTAC 852  
Qy 731 AGCAGGACTTTGGCCCTCGCAAAAGGCCAGATCGGCATCTGCTCAACGCGACTAC 790  
Db 853 AACACCTCTTTCGCGCCACACAGGAGCGGGGTGCTATCGCTTAAGCTCCATTTG 912  
Qy 791 TATGAGCCCTGGGACAGCATGAGCCTCGGGAACAGAGAGCTGTGAGCGACGATGAA 850  
Db 913 ATCAATCTCGAAGATGACTGACTATATATCAGAAAGCCAGAAAGTCTC---TTGAC 969  
Qy 851 TTTCATTGGCTGTGTTGGCAATCCCATCTTTCTTGAAGAGACTATTCAGAGAGCAT 910  
Db 970 TTGTGCTAGAGCTGTGTTGCAAAACCATATTATTATGAGCGCATCCAGAGATATG 1029  
Qy 911 AAGAGCAGCT 921  
Db 1030 AAGAACCACT 1040  
RESULT 7  
US-09-344-510B-8  
Sequence 8, Application US/09344510B  
Patent No. 6579850  
GENERAL INFORMATION:  
APPLICANT: Nabeshima, Youichi  
Kuroo, Makoto  
Sekine, Susumu  
Iida, Akihito  
TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10112-3801  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
COMPUTER: Compaq Deskpro EN  
OPERATING SYSTEM: Windows 98  
SOFTWARE: Wordpad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/344,510B  
FILING DATE: 25-Jun-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/04585  
FILING DATE: 12-DEC-1997

APPLICATION NUMBER: JP 347871  
 FILING DATE: 26-DEC-1996  
 APPLICATION NUMBER: JP 205815  
 FILING DATE: 31-JUL-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Perry, Lawrence S.  
 REGISTRATION NUMBER: 31865  
 REFERENCE/DOCKET NUMBER: 766.32  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 218-2100  
 TELEFAX: (212) 218-2200  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5032  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 ORIGINAL SOURCE:  
 ORGANISM: mouse  
 IMMEDIATE SOURCE:  
 LIBRARY: kidney  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 9..3060  
 IDENTIFICATION METHOD: E  
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-344-510B-8

Query Match 6.4%; Score 127.8; DB 4; Length 5032;  
 Best Local Similarity 53.1%; Pred. No. 1.1e-22;  
 Matches 388; Conservative 0; Mismatches 307; Indels 36; Gaps 4;

191 CCATCGCGACCAACGCGCGCAACGCGGATGCTGCTGATCACTACCGCTACGAT 250  
 364 CCGTCCCTCCCTCCCTGCTCCCTCACTGAGATGCGCAGAGATGATACAAACGCTTAC 423  
 251 GAGACTTTGATCTTTGACCAAGTACGCGCGAAAGCCTACCGCTTCCCTGTCGTG 310  
 424 CGGACACACAGAGGCGCTGCGCAACTGCGGCTCACCCACTACCGCTTCTCCATATG 483  
 311 TCGCGGATCATTCCTCCGCGCGCAGCTGATCCCGTCAACAGAGAGGAAATTGATT 370  
 484 GCGGGGTGCTCCCAATGCGACCGCGGCACTCC--CAACGCGAGGCGCTGCGCTAC 540  
 371 TACAGCAACTGATTGACGCTCTGTTGAGCGCGGCTTACGCGCTTGGGTGACTTTG 430  
 541 TACCGGCGCTGCTGAGCGCTGCGGAGCTGCGGCTGCGCGGCTGTTACCTGTAC 600  
 431 CACTGCGATGCTGCTGAGCGGCTTACGATCGCTATGAGGCTGCTCAACGTGAGAG 490  
 601 CATTGGACCTGCGACAGCGCTGCGAGACACTATGCGGATGGCGCAATGCGGCCG 660  
 491 GTCCAGCTGACTTTGAGCGGTATGCGAGTTGTGCTTTGAACGTTTGGGACCGAGTC 550  
 661 GCCGAC--ATTTCAGGATTTATGCGAGCTGCTTCGCGCACTTCGCTGTGTCAGGTC 717  
 551 CAGAACTGATCACTATCAAGACCTGATTCAGGCACTATGATATGCAACCGCG 610  
 718 AAGTACTGATCACTATCAAGACCTGATGCTGAGCTGCGACGCGGTATGCAACCG 777  
 611 AGCAACGCGCGCGCGAGGAGGAGCTTAAACAGCACTCCACCGAGGCAACACTGCC 670  
 778 CGCTGCGCGCGCGGCTGAGGAGGAGCTCCAGGCTCGG-----815  
 671 GAGCGGTGCTGCTGAGAAAGCCCAATCATAGACCAATGCGCGCGCGCTGCTTAC 730  
 816 -----GTACTGTGTGCGCACCACTACTTTTGGCTATGCAAAAGTGTGCACTCTAC 870  
 731 AGCAGGACTTTCGCGCGCGCAAAAGGCGAGATCGGACTCTGCTCAACGCGAGTAC 790  
 871 AACACCTCTTCCGCGCGCACACAGGAGGCGCGGTGTCTATGCTTAAAGTCTCCATGG 930

791 TATGAGCCCTGGGACAGCATAGACTCTGCGACAGAGCGTCTGAGCGACGATGAA 850  
 931 ATCAATCTCCAGAAATACATGACTATATATCAGAAATCCAGAAATCTC---TTGAC 987  
 851 TTTCACTTGGCTGCTTTTCCCAATCCATCTTTTGAAGAAAGCTATCCAGAGAGCATG 910  
 988 TTTGTGTAGGCTGCTTTGCGCAAAACCATATTTATGATGCGGACTACCCAGAGATATG 1047  
 911 AAGAGCACT 921  
 1048 AAGAACCACT 1058

RESULT 8  
 US-09-914-841A-2  
 ; Sequence 2, Application US/09914841A  
 ; Patent No. 6645750  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Amano Enzyme Inc.  
 ; TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE  
 ; FILE REFERENCE: 066072  
 ; CURRENT APPLICATION NUMBER: US/09/914, 841A  
 ; CURRENT FILING DATE: 2001-09-04  
 ; PRIOR APPLICATION NUMBER: P. Hei. 11-056299  
 ; PRIOR FILING DATE: 1999-03-04  
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01242  
 ; PRIOR FILING DATE: 2000-03-02  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO: 2  
 ; LENGTH: 1287  
 ; TYPE: DNA  
 ; ORGANISM: *Camellia* var. *sinensis*  
 US-09-914-841A-2

Query Match 6.3%; Score 126.2; DB 4; Length 1287;  
 Best Local Similarity 51.5%; Pred. No. 1.6e-22;  
 Matches 390; Conservative 1; Mismatches 354; Indels 12; Gaps 4;

201 CCAAGCGCGCAACGCGGATGCTGCGATCACTACACCGCTACGATGAGAGCTTTG 260  
 5 CGAATGTAGACACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 64  
 261 ATCTTTGACCAAGTACGCGCGCAAGGCTTACCGCTTCTGCTGCTGCTGCTGATCA 320  
 65 AGTGTGGAAGTTATGAGACTGATGATGATGATGATGATGATGATGATGATGATGAT 124  
 125 TACCTCGGGGAGACTTACGCGAGAGATGAAACAGAGAGTATCGCTTACCAACATG 184  
 321 TTCCCTCGCGGCGAGCTGATCCGTCACAGAGAGGATTTGATTTTACAGCAAC 380  
 185 TCATCAATGACCTTTTATGCAAGATATCAACCTTTTATTAACATCTTTTACCTGG 244  
 381 TGATTGACGCTTGTGAGCGGCTTATCAGCTTGTGGTGAATTTTATCACTGAGATC 440  
 441 TGCCCTCAGCGCTTACAGATGCTATGAGGCTGCTCAACGTTGAAGAGTCAAGCTGG 500  
 245 TTCCCAAGCCTTGAAGATGATATGAGGCTTTTAAAG--CCACACTTGTGAACG 301  
 501 ACTTGAAGCGGTATGCGAGTTGTGCTTTGAACGTTTGGGAGCGAGTCCAGAACTGGA 560  
 302 ATTTCGGGATTTTCAAGAGCTGTGCTTCAAGAGTTTGGTACCGAGTTAAACTTGG 361  
 561 TCACCATCAAGACCTCGATTCAGGCGCATCTATGATATGACACCGGACCAACGCC 620  
 362 TCACATGATGAAACCATGCTTACTCTATGGGGTTATGATGACAGTCTCCAGAC 421  
 621 CGGACAGAG--CAGACTTAAACAGCACTCCACGAGGCAACACTGCACTGAGCGGT 677  
 422 CGGCGCTGTGTTGCGTTTATGCGATTTTGGCTTAAAGGGAATCTGAGCTGAGCCT 481  
 678 GGTGCTGGAAGGCGGACATGAGCAATGCGCGCGCTGCGCTTACAGAGG 737

Db 482 ATATAGTATACCAACATTGCTTTCTCTATGCTGCTGTGTAACCTATACAGAGAGA 541  
Qy 738 ACTTGGCCCCCTGGCAAAAGGGCCAGATCGGCATCTGCTCAACGGGAGCTACTATGAGC 797  
Db 542 AATATCAGGCGATATCAAAAGGGCCAGATAGGATTAACCTATGCTATTTGATGATTC 601  
Qy 798 CTTGGGACAGCAATGAGCCTTCGGGACAAAGAGCTGCTGAGCGACGATGGAAATTTGACA 857  
Db 602 CCT---ACTCCAAATTCGAAAGCCGACAAAGATGACGACCAAGAGCCCTTGATTTGATGT 658  
Qy 858 TTGGCTGTGTTGGCAATCCATCTTTCTTGAAGAAAGCTATCCAGAGACATGAAGAGC 917  
Db 659 ATGATGTGTTATTTAGCCATTAGCTT---TGTGTAATATCCAAAAGCATGCGTAGAC 715  
Qy 918 AGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGCGGA 954  
Db 716 TCCTTGTGTAAGAGTTACCAAGCTTCACTAAGAGCA 752

## RESULT 9

US-09-122-230-8  
; Sequence 8, Application US/09122230A  
; Patent No. 5973228  
; GENERAL INFORMATION:  
; APPLICANT: Carlson, et al.  
; TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin  
; TITLE OF INVENTION: Composition in Plants  
; FILE REFERENCE: 50532  
; CURRENT APPLICATION NUMBER: US/09/122,230A  
; CURRENT FILING DATE: 1998-07-23  
; EARLIER APPLICATION NUMBER: U.S. 60/053,566  
; EARLIER FILING DATE: 1997-07-24  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1542  
; TYPE: DNA  
; ORGANISM: Pinus contorta  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1542)  
; US-09-122-230-8.

Query Match 6.3%; Score 124.8; DB 2; Length 1542;  
Best Local Similarity 54.4%; Pred. No. 4e-22;  
Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 81 TGGCCCAACGACTTGAATGGGGCTTTCGCAACGGCCGCTACAGATCGAAGCGCCGTCA 140  
Db 89 TCCCTCAGATTTCATGTTGGGACAGCCTCTTCAAGGTATCAATAGAGAGAGCAATCC 148  
Qy 141 AAGAAGTGGCCGCGGCGCCGCTTCATCTGGGACAGCTACTGCGCACTTGAGAGCAATCGGCA 200  
Db 149 GAGAAAGTGGCAAGGGTCTTAGCAGATGGGACGCTTAAACATATCTCTGTGAAATTA 208  
Qy 201 CCAAGCGCGCAACGGGATGGCTTGGCATCTACCAACCGCTAGATGAGGACTTTG 260  
Db 209 AAGTACAGACATGAGAGCTGGCACTGCAATATTCACAGATATATGAGAGATTCG 268  
Qy 261 ATCTTTGACCAAGTACGGGCGCAAGGCTACCGCTTCTCTTGTGCTGTGCGGATCA 320  
Db 269 AGCTTATGGCTTCACTTGGACTAGATGCTATATATTCTCATATCTCGGTCTGAAATCC 328  
Qy 321 TTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAAATGATTTTACAGCAAC 380  
Db 329 TTCGAGAGAGAG-----AGGTAAATTAAATGATGCTGGGATTTGAATATTAATATATC 382  
Qy 381 TGATTGACGCGCTTGTGAGGCGGGTATACAGCCTTGGGTGACTTTGATCACTGGGATC 440  
Db 383 TGAATGAGGCTTCTTGCAGAAATGGATTCAGCCGCTTGTGATCATTTTCAATTC 442  
Qy 441 TGCCTCAGGCGCTTACAGATGCTATGAGGCTGCTCAACGTGAAAGAGTCCAGCTGG 500

Db 443 TTCCCAAGACATTGAAGACTCTATGGGGATGGTAGTCTCTCAATATTA---CG 499  
Qy 501 ACTTGGAGCGGTATGAGAGGTGTGGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTGGA 560  
Db 500 ACTTGAAGCTTATGAGAGATTTGCTTCCGGGCACTTGGTACCGGTCAATATTTGGG 559  
Qy 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGACGACCAAGCCC 620  
Db 560 CGACAGTGAACGAGCCAAATCTGTTGTGCGGTGGGATACACCGTGGAAATATTTCCAC 619  
Qy 621 CGGCGAGAG 630  
Db 620 CGACAGGTG 629

## RESULT 10

US-09-122-230-6  
; Sequence 6, Application US/09122230A  
; Patent No. 5973228  
; GENERAL INFORMATION:  
; APPLICANT: Carlson, et al.  
; TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin  
; TITLE OF INVENTION: Composition in Plants  
; FILE REFERENCE: 50532  
; CURRENT APPLICATION NUMBER: US/09/122,230A  
; CURRENT FILING DATE: 1998-07-23  
; EARLIER APPLICATION NUMBER: U.S. 60/053,566  
; EARLIER FILING DATE: 1997-07-24  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1909  
; TYPE: DNA  
; ORGANISM: Pinus contorta  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (183)..(1724)  
; US-09-122-230-6.

Query Match 6.3%; Score 124.8; DB 2; Length 1909;  
Best Local Similarity 54.4%; Pred. No. 4.3e-22;  
Matches 299; Conservative 0; Mismatches 242; Indels 9; Gaps 2;

Qy 81 TGGCCCAACGACTTGAATGGGGCTTTCGCAACGGCCGCTACAGATCGAAGCGCCGTCA 140  
Db 271 TCCCTCAGATTTCATGTTGGGACAGCCTCTTCAAGGTATCAATAGAGAGAGCAATCC 330  
Qy 141 AAGAAGTGGCCGCGGCGCCGCTTCATCTGGGACAGCTACTGCGCACTTGAGAGCAATCGGCA 200  
Db 331 GAGAAAGTGGCAAGGGTCTTAGCAGATGGGACGCTTAAACATATCTCTGTGAAATTA 390  
Qy 201 CCAAGCGCGCAACGGGATGGCTTGGCATCTACCAACCGCTAGATGAGGACTTTG 260  
Db 391 AAGTACAGACATGAGAGCTGGCACTGCAATATTCACATATATATGAGAGATTCG 450  
Qy 261 ATCTTTGACCAAGTACGGGCGCAAGGCTACCGCTTCTCTTGTGCTGTGCGGATCA 320  
Db 451 AGCTTATGGCTTCACTTGGACTAGATGCTATATATTCTCATATCTCGGTCTGAAATCC 510  
Qy 321 TTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAAATGATTTTACAGCAAC 380  
Db 511 TTCGAGAGAGAG-----AGGTAAATTAAATGATGCTGGGATTTGAATATTAATATATC 564  
Qy 381 TGATTGACGCGCTTGTGAGGCGGGTATACAGCCTTGGGTGACTTTGATCACTGGGATC 440  
Db 565 TGAATGAGGCTTCTTGCAGAAATGGATTCAGCCGCTTGTGATCATTTTCAATTC 624  
Qy 441 TGCCTCAGGCGCTTACAGATGCTATGAGGCTGCTCAACGTGAAAGAGTCCAGCTGG 500  
Db 625 TTCCCAAGACATTTGAAGACTCTTATGAGGAGATGGCTGATCTCTCAATATTA---CG 681  
Qy 501 ACTTGGAGCGTATGCGAGGTGTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTGA 560

```
Db      682 ACTCGAAGCCTATGACAGATTGCTCCGGGCAATGCGTGCAGCCGTGCAAAATATGGG 741
Qy      561 TCACCATGACAGAMCCCTGATTCAGGCCATATGATATGATGACCCGCGGACGACACGCC 620
Db      742 CGACGCTACAGACGCAAAATCTGTTTGTGCGGTGGATACACCGTGGAAATATTTCCAC 801
Qy      621 CGGCGAGAG 630
Db      802 CGACGAGGTG 811

RESULT 11
US-08-688-988-9
; Sequence 9, Application US/08688988B
; Patent No. 6096545
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
; FILE REFERENCE: PPL96-03
; CURRENT APPLICATION NUMBER: US/08/688,988B
; CURRENT FILING DATE: 1996-07-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1829
; TYPE: DNA
; ORGANISM: Arabidopsis Thaliana
US-08-688-988-9

Query Match      6.0%; Score 118.6; DB 3; Length 1829;
Best Local Similarity 53.6%; Pred. No. 1.6e-20;
Matches 294; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

Qy      79  TCTGCCCAAGACCTTTGATGGGCTTCGCAACGGCCGCTACAGATCGAAG---CGC 135
Db      159  TTTCCCTGGAAGGTTTATTTATTTGGACGCGTACTGCGCATACAGATGACCAAGGTTGGA 218
Qy      136  CGTCAAGAAAGGTGGCCGCGCCGCTCCATCTGGGACACGTAATGCCACTGGAGCCATC 195
Db      219  TTTATGAAACTTGTGTGCGTGAACGACGCTTATGGACACTACTGTAGAAATATCCAGA 278
Qy      196  GCGCACCAAGCGCCGCAACGGCGATGTGGCTTGGATCACTACCAACGCTACGATGAGA 255
Db      279  GAGGTGCAATTAACGATACGCGCATGTGGCCGTTGATTTCTTCATCGTTATAAGGAAGA 338
Qy      256  CTTTGATCTCTGACCAAGTAAGCGGCGCAAGGCCCTACCGCTTCTCTGTGTGTGCGCG 315
Db      339  TATCAACTAATGAAATCTTAAACACAGACGCTTTAAGATGCTATGCGATGCGCAAG 398
Qy      316  GATCAATTCCTCGCGCGGACGCTGATCCGTCACAGAGAGGAAATTGATTTTACAG 375
Db      399  AATATTTCTCATGGGAGAAAGAAAGAAAGAGTAGTGTAACACTGTGTGCAATTTACCA 458
Qy      376  CAAACTGATTAACGCCCTGTGTGAGGCGGGGTATCACGCGTTGGGTGACTTTGTACCATG 435
Db      459  CGACCTCATTCACAGAGCTCATTAATAAATGATATACTCAATTCGTACTGTTTTCACTG 518
Qy      436  GGATCTGCTCAGGCGCTTACAGATCGCTATGAGGAGGCTGCTCAACGTAAGAGGTCCA 495
Db      519  GGAACATTCACAAATTTAGAAATGATATGCGGCGCTTTTAAAGC---GAAAGATTTGT 575
Qy      496  GCTGGACTTTGAGCGGTATGCGAGATTGTCTTGAACGTTTGGGAGCCAGATCCAGAA 555
Db      576  GAAAGATTTCCGAGAGTATGAGATTTTGTTCGAAGATAGCGTGGAAAAAGGAAACA 635
Qy      556  CTGATACCAATCAACGAMCCCTGATTCAGGCGATCTATGATGATGACACCGGAGCAA 615
Db      636  TTGATACCTTTCAATGATGAGCATGGGTTTCTCGACGCTGCTATGACGTAGGCAAAA 695
Qy      616  CGCCCGGG 624
Db      696  GGCACCTGG 704
```

```
RESULT 12
US-08-400-275-17
; Sequence 17, Application US/08400275
; Patent No. 5668295
; GENERAL INFORMATION:
; APPLICANT: Mahab, Samir Z.
; APPLICANT: Mahab, Vedpal S.
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE,
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH
; TITLE OF INVENTION: ALTERED NICOTINE CONTENT
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Ave. of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,275
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,681
; FILING DATE:
; APPLICATION NUMBER: US 07/613,160
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: PM-1696
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1545 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-400-275-17

Query Match      5.6%; Score 111.4; DB 1; Length 1545;
Best Local Similarity 49.9%; Pred. No. 1e-18;
Matches 364; Conservative 0; Mismatches 356; Indels 9; Gaps 3;

Qy      75  TAGCTGCGCAACGACTTTGAATGGGGCTTCGCAACGGCCGCTACCATGAAAGCG 134
Db      2  TGCGCTCTTTCTTAATTTCTTGTGGACAGCCTCTTCAATATTAACGATAGAAAGAG 61
Qy      135  CCGTCAAGAAAGGTGGCGCGGCCGCTCCATCTGGGACAGTACTGCACTGAGACCAT 194
Db      62  CTTTCCCTAGATGATGGAAAGCGCTCAGCAACTGGGACGTTTTTACCAATGAAGCTGTGTC 121
Qy      195  CGGCGACCAACGGCGCCACAGCGATGTGGCTTGCGATCACTACACCGCTACAGATGAG 254
Db      122  ATGTTAAGATGGAACCAAGAGATGTGGCTGTGATCACTACATCGTTATTTGGAGG 181
Qy      255  ACTTGATCTCTTGACCAAGTACAGGCGCAAGGCTACCGCTTCTGCTGTGCGTGGCG 314
Db      182  ACATCAAACTCATGACAGATATGGGTGTGAATGCTTTCGTTCTATCTATGAGGC-- 239
```





## FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...1437  
SEQUENCE DESCRIPTION: SEQ ID NO: 1268:  
US-09-107-532A-1268

Query Match 5.1%; Score 101.4; DB 4; Length 1437;  
Best Local Similarity 53.1%; Pred. No. 3.5e-16;  
Matches 263; Conservative 1; Mismatches 222; Indels 9; Gaps 2;

QY 81 TGCCCAAGACTTGAATGGGCTTCCCAAGCGCGCTCCAGTGAAGGCGCGTCA 140  
DB 26 TTCCAAAGATTCTGTGGGGCTCTCTCTGAGGCTTCACTGAAAGAGCTTGGC 85  
QY 141 AAGAGGTGGCGCGCGCGCTCCATCTGGGACAGTACCTGCACTGAGCCATCGGCA 200  
DB 86 AAGAGATGGAGAGAGAGATCAGTCTGGGATTCATTTGACGATACCTGAGAAAGACAT 145  
QY 201 CCACGGCGCCACGCGCATGTGGCTTGCATCACTACACCGCTACGATGAGACTTGG 260  
DB 146 TTAAAGAACGAAACGATGATTTGGCAGTAGACCACTATCATCGTTATPAAAGAGATATG 205  
QY 261 ATCTCTTACCAAGTACGCGCGCAAGCGCTACCGCTTCTCTTCTGCTGCTGCGGATCA 320  
DB 206 CTTTATATGAAGACAGAGCTCTGAAAGCTTACCGCTTCTCTGATGCTTGGACACGATTT 265  
QY 321 TTCCCTCGCGCGCGCTGATCCGTCACAGAGAGGAAATTGATTTTACAGCAAC 380  
DB 266 TTCCCTTAAG-----GACGAGAGAAAGTCAATCAAGCTGGGCTGCTTTTATGAGAGAT 319  
QY 381 TGATGAGCGCTTGTGAGGCGGGGTATCAAGCCCTTGGGTGACTTGTACCACTGGGATC 440  
DB 320 TGATGACCAATGATGATGAAACGATATGAAACCTATGTTGACTTATATCATTTGGATC 379  
QY 441 TGCGTCAGCGCTTACAGATCGCTATGAGGCTGCTCAACGTGAGAAAGTCCAGCTGG 500  
DB 380 TTCCACAGATTCGAAAGATGATATGATGTTG---GAATCAGCTGAGATCATTTGAAG 436  
QY 501 ACTTGGAGGCTATGCGAGGCTGTGCTTGAACGTTTGGGACCGAGTCCGAACTGSA 560  
DB 437 ATTTCACCTATATGAGAAACATATTTGAGGCAATTCGCGGAGAAAGTCACTTACTGGA 496  
QY 561 TCACCATCAACGAC 575  
DB 497 TTACTTGAATGAC 511

## RESULT 15

US-08-232-463-14  
Sequence 14, Application US/08232463

## GENERAL INFORMATION:

PATENT NO. 5670367

## APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

## TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-9300

TELEFAX: (703) 683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PTZ9PT-F15

US-08-232-463-14

Query Match 5.0%; Score 100; DB 1; Length 7218;  
Best Local Similarity 4.5%; Pred. No. 1.5e-15;  
Matches 13; Conservative 209; Mismatches 64; Indels 0; Gaps 0;

QY 1528 TTCTCTTGCATTCAGCTCTATGATCTTCTCTCTTTTCCCTCCCTCTGT 1587  
DB 1174 YY 1233  
QY 1588 CCCTCTCTCTACCTCTCATATTCCTCTATACCCCGCTCTCTCATGACCCATG 1647  
DB 1234 YY 1293  
QY 1648 CTCTGCGCTTGCGCCCTCTCTGCAATTCGCTCTTATCACTTATGCTGT 1707  
DB 1294 YY 1353  
QY 1708 TTACTGCTTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1767  
DB 1354 YY 1413  
QY 1768 TATCTGCGCTTGCGCTCATTTGGCAACGATACATGACAAATTTCAA 1813  
DB 1414 YY 1459

Search completed: March 28, 2004, 20:56:53  
Job time : 153.443 secs



181 CCACCTGGAGCCATGCGCCACCAAGCGCCCAAGCGCGATGAGCTTGCACTACCA 240  
181 CCACCTGGAGCCATGCGCCACCAAGCGCCCAAGCGCGATGAGCTTGCACTACCA 240  
241 CCGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGTTTC 300  
241 CCGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGTTTC 300  
301 CTGTGCGATGCGCGGATCATTTCCCTGCGCGGAGCGGTGATCCCGTCAAGAGAGG 360  
301 CTGTGCGATGCGCGGATCATTTCCCTGCGCGGAGCGGTGATCCCGTCAAGAGAGG 360  
361 AATTGAGTTTACAGCAAACTGATTTGACCGCTTTGAGCGCGGATACAGCGCTTGGGT 420  
361 AATTGAGTTTACAGCAAACTGATTTGACCGCTTTGAGCGCGGATACAGCGCTTGGGT 420  
421 GACTTTGACCACTGGGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCA 480  
421 GACTTTGACCACTGGGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTCA 480  
481 CGTGGAGAGGTCAGCTGGAATTGAGCGGATGAGCGGATGAGCTTTGAGCGTTTGG 540  
481 CGTGGAGAGGTCAGCTGGAATTGAGCGGATGAGCGGATGAGCTTTGAGCGTTTGG 540  
541 GAGCGAGTTCAGAACTGATCACCATCAAGAACGCTGATGAGCGCATCTATGAGATA 600  
541 GAGCGAGTTCAGAACTGATCACCATCAAGAACGCTGATGAGCGCATCTATGAGATA 600  
601 TGCCACCGGAGCAAGCGCGCGGAGAGAGCAAGATTAAGAGCACTTCAAGAGGCGAA 660  
601 TGCCACCGGAGCAAGCGCGCGGAGAGAGCAAGATTAAGAGCACTTCAAGAGGCGAA 660  
661 CACTGCACTAGGCGGTGCTGCTGAGAAAGCGCAAGATCATGAGCAATGCGCGCGGT 720  
661 CACTGCACTAGGCGGTGCTGCTGAGAAAGCGCAAGATCATGAGCAATGCGCGCGGT 720  
721 GCGCGTCTACAGCAGGAGCTTTGCGCCCTGCAAAAGGCGCAATGCGCATCTGCTCAA 780  
721 GCGCGTCTACAGCAGGAGCTTTGCGCCCTGCAAAAGGCGCAATGCGCATCTGCTCAA 780  
781 CCGGCACTATGAGCGCTGCGGAGCAAGTACGAGCGCTGCGGAGAGGCTGCTGAGCG 840  
781 CCGGCACTATGAGCGCTGCGGAGCAAGTACGAGCGCTGCGGAGAGGCTGCTGAGCG 840  
841 ACGGATGGAATTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
841 ACGGATGGAATTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
901 AGAGAGCATGAG 960  
901 AGAGAGCATGAG 960  
961 CATGCTCAATGCGGAG 1020  
961 CATGCTCAATGCGGAG 1020  
1021 GCGGCACTAG 1080  
1021 GCGGCACTAG 1080  
1081 GAGAGATTAAG 1140  
1081 GAGAGATTAAG 1140  
1141 CCGGAGCATGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1141 CCGGAGCATGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1201 CATACCGAG 1260  
1201 CATACCGAG 1260  
1261 CAAAG 1320

1261 CAAAG 1320  
1321 TACCAG 1380  
1321 TACCAG 1380  
1381 ATGCTGAGATGAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1381 ATGCTGAGATGAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1441 GCGGAG 1500  
1441 GCGGAG 1500  
1501 TAAAGTGGCGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1501 TAAAGTGGCGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1561 TCTCTCTTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1620  
1561 TCTCTCTTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1620  
1621 CCGCGCGCTTCTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1621 CCGCGCGCTTCTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1681 GCGCTTATGAGGCTTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
1681 GCGCTTATGAGGCTTATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
1741 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
1741 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
1801 AGCAGAGGTTCAAG 1860  
1801 AGCAGAGGTTCAAG 1860  
1861 AAGGCGCATGTTTACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1920  
1861 AAGGCGCATGTTTACCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1920  
1921 CAG 1980  
1921 CAG 1980  
1981 AAAAAAAAAA 1991  
1981 AAAAAAAAAA 1991

RESULT 2  
US-10-026-140-3  
; Sequence 3, Application US/10026140  
; Publication No. US20030114330A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuurt, Frits  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: BG5 Beta-Glucosidase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC697  
; CURRENT APPLICATION NUMBER: US/10/026,140  
; CURRENT FILING DATE: 2002-03-26  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1455  
; TYPE: DNA  
; ORGANISM: Trichoderma reesei  
US-10-026-140-3

Query Match 73.0%; Score 1454.2; DB 14; Length 1455;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTGGCTAGCTCTGCCCAAGCATTTGATGGAGCTTCCGCAAGCGCGCTTAC 121  
 DB 1 ATGCCGAGTGGCTAGCTCTGCCCAAGCATTTGATGGAGCTTCCGCAAGCGCGCTTAC 60

QY 122 CAGATGAAAGGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCCATCTGGGACAGTACTGC 181  
 DB 61 CAGATGAAAGGCGCGCTCAAGAAAGTGGCGCGCGCGCTTCCATCTGGGACAGTACTGC 120

QY 182 CACCTGAGGCGATCGGCGCAACGCGCGCAAGCGGAGTGGCTTGCATCTACAC 241  
 DB 121 CACCTGAGGCGATCGGCGCAACGCGCGCAAGCGGAGTGGCTTGCATCTACAC 180

QY 242 CGCTACGATGAGACTTTGATCTTGTACCAAGTACGCGCAAGGCTTCTCC 301  
 DB 181 CGCTACGATGAGACTTTGATCTTGTACCAAGTACGCGCAAGGCTTCTCC 240

QY 302 TTGTCTGTGTCGCGGATCATTTCCCTGCGCGGAGGCTGATCCCGTCAAGAGAGGA 361  
 DB 241 TTGTCTGTGTCGCGGATCATTTCCCTGCGCGGAGGCTGATCCCGTCAAGAGAGGA 300

QY 362 ATTGAGTTTACAGCAAACTGATGAAGCGGCTGTTGAGGCGGATCAAGCGCTGGGTG 421  
 DB 301 ATTGAGTTTACAGCAAACTGATGAAGCGGCTGTTGAGGCGGATCAAGCGCTGGGTG 360

QY 422 ACTTGTACCACTGGGATCTGCTCAGCGGCTTCAAGCATGCTTACGAGGCTGCTAC 481  
 DB 361 ACTTGTACCACTGGGATCTGCTCAGCGGCTTCAAGCATGCTTACGAGGCTGCTAC 420

QY 482 GTGGAAAGGTCCAGCTGACCTTTGAGCGGATGCGAGGTTGTGCTTTGAAGCTTTGGG 541  
 DB 421 GTGGAAAGGTCCAGCTGACCTTTGAGCGGATGCGAGGTTGTGCTTTGAAGCTTTGGG 480

QY 542 GACCGAGTCCAGAACTGGATCACCATCAAGACCCGATGAGGCTTACGAGGCTTACGAT 601  
 DB 481 GACCGAGTCCAGAACTGGATCACCATCAAGACCCGATGAGGCTTACGAGGCTTACGAT 540

QY 602 GCCACCGGCGAGCAAGCGCGCGCGGAGGAGCATTTAAAGAGACTTCCACGAGGCGAAC 661  
 DB 541 GCCACCGGCGAGCAAGCGCGCGCGGAGGAGCATTTAAAGAGACTTCCACGAGGCGAAC 600

QY 662 ACTGCCACTGAGCGCTGCTGCTGGAAGGCCCAAGTCAATGAGCATGCGCGCGCTG 721  
 DB 601 ACTGCCACTGAGCGCTGCTGCTGGAAGGCCCAAGTCAATGAGCATGCGCGCGCTG 660

QY 722 GCGGCTACAGGAGGACTTTCGCGCGCGGCAAGAGGCGGAGTCCGAGCTTGGCTCAAC 781  
 DB 661 GCGGCTACAGGAGGACTTTCGCGCGCGGCAAGAGGCGGAGTCCGAGCTTGGCTCAAC 720

QY 782 GCGGACTACTATGAGCCCTGGGACAGCAATGAGCTTGGGACAAAGAGGCTGTGAGCA 841  
 DB 721 GCGGACTACTATGAGCCCTGGGACAGCAATGAGCTTGGGACAAAGAGGCTGTGAGCA 780

QY 842 CGGATGGAATTTCACTTGGCTGTTTGCATTCCTTTTGAAGAGACTATCA 901  
 DB 781 CGGATGGAATTTCACTTGGCTGTTTGCATTCCTTTTGAAGAGACTATCA 840

QY 902 GAGAGATGAGAAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCC 961  
 DB 841 GAGAGATGAGAAAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCC 900

QY 962 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAACATCCAGTTGCGG 1021  
 DB 901 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAACATCCAGTTGCGG 960

QY 1022 CGGCACCTAAGCGGTCCGTCGCCGAGAGCGAGCTATCTCGGCGCATTCATGAGCACAG 1081  
 DB 961 CGGCACCTAAGCGGTCCGTCGCCGAGAGCGAGCTATCTCGGCGCATTCATGAGCACAG 1020

QY 1082 GAGAAATAGAGACGCGCCCGCTTGGCGAGAGAGCGGCTGCTGCTGCTCTGC 1141  
 DB 1021 GAGAAATAGAGACGCGCCCGCTTGGCGAGAGAGCGGCTGCTGCTGCTCTGC 1080

QY 1142 CCGGACATGTTCCGGAAGCATCTCGCGGCTGTACGCGCTGTACCGCAAGCCATCTAC 1201  
 DB 1081 CCGGACATGTTCCGGAAGCATCTCGCGGCTGTACGCGCTGTACCGCAAGCCATCTAC 1140

QY 1202 ATCAGCGAAGAGGAGGCGCGCTTGGAGAGGAGAAATGACGTCGCGAGGCGCTC 1261  
 DB 1141 ATCAGCGAAGAGGAGGCGCGCTTGGAGAGGAGAAATGACGTCGCGAGGCGCTC 1200

QY 1262 AACGACCCCTTCCGATCCGCTTCTTGAATGCACTGCACTGATTTTCAAGGCGATT 1321  
 DB 1201 AACGACCCCTTCCGATCCGCTTCTTGAATGCACTGCACTGATTTTCAAGGCGATT 1260

QY 1322 ACCGAGAGCGGCTGCTGCTCAAGGAGTACTTTCGCTGCGCTTCTGATTAATTGAA 1381  
 DB 1261 ACCGAGAGCGGCTGCTGCTCAAGGAGTACTTTCGCTGCGCTTCTGATTAATTGAA 1320

QY 1382 TGCTCAGATGCTACGAGACCCAGATTCGCGCTCAGCTTCAAGATCAACACCTCAAG 1441  
 DB 1321 TGCTCAGATGCTACGAGACCCAGATTCGCGCTCAGCTTCAAGATCAACACCTCAAG 1380

QY 1442 CGCAGCCCAAGAGTCTGCGCTGCTCTCAAGAGATGTTTGGCGCGCAGAGGTT 1501  
 DB 1381 CGCAGCCCAAGAGTCTGCGCTGCTCTCAAGAGATGTTTGGCGCGCAGAGGTT 1440

QY 1502 AAGTGGCGGATTA 1516  
 DB 1441 AAGTGGCGGATTA 1455

RESULT 3  
 US-10-369-493-37092  
 ; Sequence 37092, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; PRIOR FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; NUMBER OF SEQ. ID NOS: 47374  
 ; SEQ. ID NO 37092  
 ; LENGTH: 1470  
 ; TYPE: DNA  
 ; ORGANISM: *Aspergillus nidulans*  
 US-10-369-493-37092

Query Match 16.0%; Score 317.8; DB 15; Length 1470;  
 Best Local Similarity 53.8%; Pred. No. 2,6e-83;  
 Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

QY 81 TGCCCAAGCACTTGAATGGGCTTGGCAAGCGCGCTTACGAGTCAAGAGGCGCTCA 140  
 DB 2 TGCCCAATGACTTCTTCCATGATGATGCAAGCGCGCGCTTGAAGAGGCGCTGCA 61

QY 141 AAGAGGTGCGCGCGCGCTTCCATCTGGGACAGTACTGCACTTGAGGCACTCGCGCA 200  
 DB 62 ACAAGATGCAAGAGGCTTCTGATCTGGGACACATTCGCGCATCTCCAGCAAGTCA 121

QY 201 CCAAGCGCGCAAGCGGATGCTGCTGCTGATCACTCAACCGCTACGATGAGAGCTTTG 260  
 DB 122 AAGACAAATAGCAATGGGATGAGCGCTGAGGATTTTACCGGAGAGATGTTG 181

OY	261	ATGCTTGACCAAGTACGGGGCAAAAGGCTACAGCGCTTCTCTGTGTGGGTGGCGGATCA	320
Db	182	CCTGATGAAGTGTATGTGTGTCAAGCGTTATGATTTCTGCTATCTGGTCCGGATTA	241
OY	321	TTCCCTCGGCGGAGGCTGGATCCGTCACAGAGAGGAAATTGATTTACAGCAAC	380
Db	242	TCCCACTCGGAGGGCGCGACGACCCGGTCAATAGCAAGGAACTCAAGTACTACAGAGCC	301
OY	381	TGATTTAGCGCTGTBAGGCGGGGTATCAAGCTTGCGGTGGATCTTTGTATCACTGGGATC	440
Db	302	TGTAGACAGAGTACTCAATATATGGATTATGCGCATTCGTAAACCCTCTTCACTGGGATG	361
OY	441	TGCTCAAGCGCTTACAGATGCTATTGAGAGCTGGCTCAACGTGAAAGGTCACAGCTGG	500
Db	362	TTCCGAGAGGCTTTGAGAGATCGGTACGGGGCAATGCTGAACCAAGAGAGGTTCACTTCG	421
OY	501	ACTTTGAGCGGTATGCCAGGTGTGTCTTTAAAGCTTTTGGAGACCAAGTCCAAATCTGAG	560
Db	422	ACTTGTCCGTACGGCGGGTGTCTGTTTCGAACCCCTAGGAGCCAAAGGTTCCGTCACTGAG	481
OY	561	TCACCATCAAGAACCCCTGAGATTACAGCCATCTATGTGATATGCTCACCGGACAGACGCC	620
Db	482	TTACATTCATGAACCTGTGTGTATCTGCTGGAGGTTATGGCGCAGGTTCATGCTC	541
OY	621	CGGCGAGAGCAGCATTTAACAGCATCTCACCGAGGGCACACTGSCCATCTGAGCCGTGGC	680
Db	542	CTGCACGGTCACTGTTTTCGCAACCTCAATGAAAGGGGTGATTCCTTCACAGAACCCGTCA	601
OY	681	TGCGTGGAAAGGCCAGATCATGAGCCATGCGCCGCGCGTGGCCGTCTAAGCAGGAGCT	740
Db	602	TGCTGGGCAACAAACTGTGTGACCAAGGCCATGTTTCCAACTTTACAGAGAGGTGT	661
OY	741	TTGCGCCCTCGCAAAAGGGCCAGATCGGCACTCTGCTCAACGGGCACTACTATAGGCTCT	800
Db	662	TCCAAACCCAGAGAAAGAAACAATTGGCATCACTCTCAATGAGCACTGGTCGGAACTTT	721
OY	801	GGGACAGCAATGAGGCTTGGGGCAAGAGGCTGTGAGCCGACCGATGAAATTTCACTTGG	860
Db	722	GGGATGAAGACGATTCGGGGAGCCAGAGACGACGAAACGGGCCAGGGAAATTCAGATTCG	781
OY	861	GCTGTTTTCCAATCCCATCTTCTTTGAAGAGAGCTATCCAGAGAGCATGAAAGAGCAGC	920
Db	782	CTTGTTCCCGGACCCCTTATACAAAGACAGCGACTACCCGAGCTCAATGCGGGCTCAAC	841
OY	921	TGGGGAAGAGGTTTCCAGCCCTCACTCCCGGGAGCTTTGCCATCTCAATGCCGGAGAGA	980
Db	842	TGGGGGACCGTTTACCGGTTTCACTCCGAGAGGT---CAAACTGGTACTTGAAGATT	898
OY	981	CCGACTTCTACGGCATGTAATTACTACACATCCCAAGTTGCGCGGCACTTACAGCGTCCG	1040
Db	899	CAGATTTTACGGTATGAACCTGTATCAAGACTTTTGTGCAGACAAAGATACGCTC	958
OY	1041	TCCCGAGACGGACTATCTCGGCGCATTCATGACACACAGAGAAATTAAGACCGCAGCC	1100
Db	959	CAGATATCATATACCAAAAGGCAATGTCAATGTTCACAGATACAAACGCAAGAGCGTCT	1018
OY	1101	CGGTTGGCAGAGAGAGCGGCTCGGCTGGCTGCGCTCTCTGCGCGGACATGTTCCGAGAGC	1160
Db	1019	CTCGAGGCGAAGATCCGATACGCTTGGCTTCCGACCGGACCCACTGGATGGGCGAAGC	1078
OY	1161	ATCTCGCCCGGGGTATCGGCTGTACCGGCAAGCCCATCTTACATCACTCGAAGCGAGTGC	1220
Db	1079	TGCTCAATTGATTGGAAACCATATCATGTGCTTATATGTCCACGAAAAATG---GCA	1135
OY	1221	CGTGCCCTGAGAGAGACATGAGCGTGGAGAGAGCGGTCAAGACCCCTTCCGATCC	1280
Db	1136	CGACAGCAAAAGAGAGACAGCTCCACACCAAGAGTACTCATCACACTTTCGTATGC	1195
OY	1281	GATATCTTGACTCGCACTTGTGACTCGATTTTCCAAAGCCATTACCCAGAGCGGCGTGTGC	1340
Db	1196	GATTTCTTGAAGGCTATGTTGGATGTGGCCCGTGTGTCAAAAGAGATGGAAGTGATA	1255
OY	1341	TCAGAGGATATCTTGGGTGGGCGTTCGTGATATCTGG	1379

Qy	645	ACTCCACGAGGGGAAACATGCGACTGAGAGCGGTGGCTGGTGGAAAGGCCACGATCATGA	704
Db	687	-----TTGGGCGGGCACTTCGCCACCGAGCCGTATACGCGCGGCAACACTCTCTCTCG	741
Qy	705	GCCATGCCCCGCGCGTGGCCGCTCTACAGCAGGGAGCTTTCGCCCCCTCGCAAAAGGGCCAGA	764
Db	742	CGCAGCGTTCGCGCTGTCAAGCTGTACCCGACGAGTACAAAGTTTCGTACGCGCGGCAAGA	801
Qy	765	TGCGATCTCGCTCAAGGGCCACTCTATGAGCCCTGGGACAGCAATAGAGCTCGGGACA	824
Db	802	TTGGGATCTCTGAGACTTCCTGGTGGTACGAAACC---GCTGACCAAGTCCGTGAGAGACG	858
Qy	825	AGAGAGCGTGTGAGGAGCGAGTGAATTTCACTTGGCTGGTGGTTGCAATCCCATCTTCT	884
Db	859	AGTACGCGCGGACCGGGCTCGGATGTTTCACCTCGTGGTCTGTTCTGTGACACCCCATCACC	918
Qy	885	TGAAGAAAGCACTATCCAGAGAGCATGAAGAAAGCACTGGGCGAGAGGCTTTCAGCCCTCA	944
Db	919	---ACGGCCACTACCCGGAGAGAGTGAAGAAATGTCATGAGGGAGGCTGCCCACTTCA	975
Qy	945	CTCCCGGCGGACTTTGGCCATCTCCATATGCCGAGACCGGACTTCTACCGCATGAATTTCT	1000
Db	976	CTTTGACAGCTGTGCAATGGCTAAAGGCTCAGCG--GACTACGTTCGCACTAACACT	1033
Qy	1005	ACACATCCCAAGTTCGCGCGCCACTAGACGATGCCCTGCCCGCCAGACGCACTATCTCGGG	1066
Db	1033	ACACACAGTACTAGCCGACGCAACTTCGTACACGCCACAGAGACCAACTACCCGACAGATT	1099
Qy	1065	CCATCCATGAGCACCAGAGAAATTAAGACGCGAGCGCCGCTGGTGCAGAGAGAGCGGCTTCG	1122
Db	1093	GGAATGCAAAATTTCTATGAGCGAGATGGTGTGCCCATTTGGCAAAAGGCGTACTCGG	1155
Qy	1125	CCTGGCTGCGCTCCGCGCCGACATGTTCCGGAACATCTGCGCGGGGTGTAAGGCGCTGT	1188
Db	1153	ACTGGCTTTACGTGCTTCATGAGGGGCGCTTACAGAGCTTGTATTTGGACCAAGAGAAAT	1211
Qy	1185	ACGGCAAGCC---CATTTACATCACCGAAGACGATGCCCGTGCCTCGGAGAGAGAACCA	1244
Db	1213	TCAACAGCCCTGTGATGCTCATCGAGAGAAACGAATTTGACCAAGCCTGGAAATGAGACT	1277
Qy	1242	TGAGCTGGAGAGAGCGCGTCAACGACCCCTTCCGATCCGATACCTTTGACTGCACTTGG	1300
Db	1273	TGCGCTTC-----GCTCTGTACGCAAGTTTACAGATGACATCTTCGAAAGTACCTGT	1322
Qy	1302	ACTCGATTTCCAAAGGCCAATTACCCACGAGCGGCGTGTGCTCAAGGAGGTACTTTGCTGGG	1361
Db	1327	ACGAGCT---CCAGTGGCGCATACCGAGAGGTGCAAAAGTCTTTCGGCTACCTTCGGGTGGT	1388
Qy	1362	CGTTGCTCGATTACTTGGAAATGTCAGATGGCTACGAGACCCAGATTCGGCGTCAAGTTCA	1421
Db	1384	CGGTGCTGGAACAATTGAGTGGCGGCTCGGCTTCACTCCAAAGTTTGGAGTCGTGACG	1441
Qy	1422	CAGACTACCAACCACTCAAGCGCACGCCCAAGAAGTGTGCTGCTGCTCAAGGA	1476
Db	1444	TGAGACCGGAAACGTTTCGTCCGTAACCCCAAGAGACTCAGCTGCTGTTCAGAA	1498

RESULT 5  
 US-10-425-114-19747  
 ; Sequence 19747, Application US/10425114  
 ; Publication No. US2004003488A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28

```

; NUMBER OF SEQ_ID NOS: 73128
; SEQ_ID NO 19747
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3151-012-C9_FLI
US-10-425-114-19747

```

Query Match	11.6%;	Score 230.4;	DB 12;	Length 1794;
Best Local Similarity	52.3%;	Pred. No. 2.4e-57;		
Matches 722; Conservative	0;	Mismatches 616;	Indels 42;	Gaps 8;

OY	101	GGCTTTCGACAGGGCGCGCTTACCAATCGAAGAGCCCGTCAAGAAAGAGTGGCGGCGCCG	160
Db	237	GGCAACCGCACCTCTCGTATCCAAATGGAAGCGCGTACCTCGAGGGGCAACAAAGCTTGG	296
OY	161	TTCAATCTGGGACACAGTACTGCCACCTGGAGCCATCGCGCACCAACGGCGCCACAGGCGAT	220
Db	297	AGCAACTGGGATGTCTTTCACGCCACAGTACCAAGGAAGAAATCGAAGACGGAAAGCACCGGAAAT	356
OY	221	GTGGCTTGGCATCACTACACACCGGTCAGATGAGGACTTTGATCTCTTGACCAAGTACGGC	280
Db	357	ATCGCGGATGATCACTATCATCTGTTACAGGAGTGAACATTGAGCTCATGCACTTCTCGGC	416
OY	281	GCAAAGGCTTACGCGCTTCTCCTTGTGCTGTGCGGATCATTTCCCTTCGCGGACAGCTG	340
Db	417	ACCAACGCGCTACAAATTTTCTATATCATATGGCCAGAAATTTCTTCC---GAGAGGGAGATTCC	473
OY	341	GATCCCGTCAACGAGAGGAGAAATTGAGTTTATACAGCAACATGATTTGACCCCTGTGAGG	400
Db	474	GGCAGGTCAAACCAAGACAGATGCGCTTCTACAAACAGACTCATGATCGCTCCTCTGTC	533
OY	401	CGGGGTATCAACGCGCTTGGGTGACTTTGTATCCACTCGGATCTGCGCTCAGGCGCTTACGAT	460
Db	534	AAAGGATAGAGCGGTTCTGTGACGCTGTCCCACTACGACATCCGCAAGAGCTGGAGAC	593
OY	461	CGCTATGAGGCTGGCTTCAACGTGGAAGAGGTCCAGCTGGACTTTGAGCGGTATGCGAG	520
Db	594	AGGTAACGGGGGTTGAGCTCGCGGGGCGGAGGCGGGCGGACTTCGCGCACCTGGCGGAC	653
OY	521	TTGTGCTTTTGAACCGTTTGGGAGCCGAGTCCAGAACTGAGTACACATCAACGACGCCCTGG	580
Db	654	GTGTGCTTTCGGGCGGCTTCGGGACCGGCTCAGGCACTGTGGCGCACTTTCAAGAGCCCAAC	713
OY	581	ATTGAGGCGCATCTATGATATGCCACCGGACGACACCGCCCGGCGAGGAGCAGCATTTAAC	640
Db	714	GTGCGGGTCAACCAAGGGGCTACATGCTGGGACCTAACCGCGGGGCGAGTGC-----TCG	767
OY	641	AAGCATCTCAACCGAGGGGACACATCTGCCATGAGCCGTGAGCTCTGTGGAAAGCCCAATC	700
Db	768	CGCTCTCGCGCCGAGGCACTTCGAGCGCCGAGGCTTACGTGCGCGGCAACAAAGTCTGC	827
OY	701	ATGAGCCATGCGCCCGCGCTGCGGCGGTCTACAGCAGGAGCTTTCGCCCTCCGCAAAAGGCG	760
Db	828	CTGGGCCACGCGCGCGCGCTCCCAATCTACAGA-----CCAATACCAAGAGGCG	878
OY	761	CAGATCGGCATCTCTGCTCAACGCGGACTACTATGACCCTTGGAGCAAGCAATAGAGCTTCG	820
Db	879	TTGATTCGGCATCTGATATGTCCACTGTGTGTTCTGTGCGCTG---ACGGAAGCGCCGCGC	935
OY	821	GACAAGGAGGCTGCTGAGCGACCGATGGAATTTTCAACTGGCTGTGTTGCCAATCCCATC	880
Db	936	GACCGGCTGGGACATGAGCGGAGCGCTGGCTTGCACGTTCCATATGTTTTCTTGACCCGAT	995
OY	881	TTCTTGAAGAAGACTATCCAGAGAGCATGAAGAAGACAGCTGGGCGAGAGGCTTTCAGCC	940
Db	996	ATCT---ACGGAAGACTACCTCCGGAATGCGCGGCTCTCGGAGCTCAGGCTCGCGAC	1052
OY	941	CTCACTCCCGGGCACTTTGGCATCTCTCAATGCGCGGAGACCGGACTTCAACGGATGAAT	1000
Db	1053	TTCTCGCGCGGAGATGAGAGGACCT---GGGCTACGCGGCTGAGACTTCACTATCGGATCAAC	1109

QY 1001 TACTACATCCAGTTGCG-----GCGCACCTAGACGATCCCGTCCCGAG 1048  
DB 1110 CACTACGACGCGCTGACGCGCAGGAGATGTCATTTCTCGCAGATTATGCCGTCGGGG 1169  
QY 1049 ACGGACTATCTCGGCGCCATTCATGACACACGAGAAATAGACGCGACGCCCTTGGC 1108  
DB 1170 CAGGAGTTCCACCAATCGCTGCGGCCCTACACGCGAGAAAGAGAGGATTTCCATCGGG 1229  
QY 1109 GAGGAGCGCGCCCTCGCTGCGCTCGCTCGGCCCGGACATGTTCCGGAAGCATCTCGCC 1168  
DB 1230 CCTCGACAGAGATGCGGACGCTTCTACGTGTTCTGACGGGATAGAGAAATGCTCAC 1289  
QY 1169 CGGCTGTAACGCGCTGACGCGCA---GCCCATTAATCAACCGAGAAAGCATGCCGTGC 1225  
DB 1290 TACATCATGAACAGATACAGCAATCTCCCATGTTCAATACCGAAATGCTACGACAA 1349  
QY 1226 CTTGAGAGAGAAATGATGCTGCGAGAGAGCCGTCACAGCCCTTCCGATCCGATAC 1285  
DB 1350 GGTGGGATGTTATTAATAACCGGCGAGGATGCGCTTGACGACGAGGCGAGATACAG 1409  
QY 1286 TTTGACTGCACTTGGAATTCGATTTCCAGGCGCATTAACCGAGAGCGGCTGTCGTAAG 1345  
DB 1410 TACTCGAGGCTTACCTGCGAAAACCTGCGCAAGTCATCACTGACGCGCGCATGTCGCT 1469  
QY 1346 GGGTACTTGGCGTGGCGTGGCTGATTAATTTGGAATGTCAGATGCTAGAGCCGAGA 1405  
DB 1470 GCGTACTTATCTGCTCTCTCATGCAATTTGAGTGGCTATATGAGTACACCTTCGCG 1529  
QY 1406 TTTGGCGTCAAGCTTCAAGACACGACCCCTCAAGCGCGACCGCAAGAGTCTGCGCTG 1465  
DB 1530 TTGGCGCTTCACTACGTTGATCAACAAACGACGAGAGAAAGCCCAATCGTGGCGCTG 1589

## RESULT 6

US-10-425-114-30939

Sequence 30939, Application US/10425114

Publication No. US20040034888A1

GENERAL INFORMATION:

APPLICANT: Liu, Jindong

APPLICANT: Zhou, Yihua

APPLICANT: Kovalic, David K.

APPLICANT: Screen, Steven E

APPLICANT: Tabaska, Jack E

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53313) B

CURRENT APPLICATION NUMBER: US/10/425,114

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 73128

SEQ ID NO 30939

LENGTH: 1781

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01\_FLI

US-10-425-114-30939

Query Match 11.4%; Score 227.4; DB 12; Length 1781;

Best Local Similarity 52.2%; Pred. No. 1,9e-56;

Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

QY 81 TGCCCAAGCACTTGAATGAGGGCTTGCACAGCGCGCTTACAGATGAGAGCGCGTCA 140  
DB 230 TCCCAAGGGGTTGTGTTGCGGACGCGAGCTGCGGTCACAGTGTGAAGGGCGCGT 289  
QY 141 AAGAAGTGGCGCGCGCGCTTCATCTGGGACACGTAAGTGCACCTGAGCCATCGCGCA 200  
DB 290 CCACCAAGCGCGCGCGCGCTTCATCTGGGATTCATTCGCGACGTCGCCAAGAAATTTG 349  
QY 201 CCAAGCGCGCAAGCGGATGCTGCGATCACTCAACCGGTCAGATGAGGATTTG 260  
DB 350 CAGGAATCAAAATGAGACGTTGCAATGATCAATACATCGCTTCAAGAGACCTCG 409

QY 261 ATCTTTGACCAAGTACGCGCGCAAGGCTTACGCGCTTCTCTGTGCTGCTGCGGATCA 320  
DB 410 ATCTATATAAAGTTTGAATCTTTGATGCTTACCGGTTCTCAATCTCATGTGCTCCAGATCT 469  
QY 321 TTCCCTCGCGGAGCGGTGATCCGTCACGAGAGGAAATGAGTTTATACGCAAC 380  
DB 470 TCCCGAATGGGAAAGG-----AAAGTCATCCAGAAAGGTGAGCGTATTAATTAATTT 523  
QY 381 TGATTAAGCGCTGTTGAGGCGGGGTATCAAGCGCTTGGGTGACTTTTATCACTGAGATC 440  
DB 524 TGATTAATATCTGCTTGAAGAGGATGATCTCTTATCATCACTTTACCATATATGTC 583  
QY 441 TGCCTCAGCGCTTACCATGCTATGAGGCTGCTCAACGTGAGAGAGGTCCAGCTGCG 500  
DB 584 TTCTCTTTCGCTTGAAGAAATATGAGGCTGTTAAGCGCGAAGATGCGGACTTGG 642  
QY 501 ACTTTGAGCGGTATGCGAGGTTGCTTGAACGTTTGGGAGCCGATCCAGAACTGGA 560  
DB 643 --TTTACAGACTATGCTGACTTCTGTTTAAGACCTACGCGGATGCGTTAAAGCACTGGT 700  
QY 561 TCACCATCAAGACGCGCTGATTTACGCGCATCTATGATATATGCCACCGCAGCAAGCGCC 620  
DB 701 TTACATTAATAGGCGCAAGATATAGCGCTACTTGGCTATGACAGGCGTCAATCTCTC 760  
QY 621 CCGGCGAGGACGACATTAACAAGCACTCCACGAGGCGAACCTGCCACTGAGCGGTGGC 680  
DB 761 CTAAAGGTC-----ACGATGCGCTGCTGCTGCGGAAATTCAGCAACCGAAGCTTACA 814  
QY 681 TCGCTGAAAAGCGCCAGATCATGAGCATGCGCGCGCGCTGCGCTTACAGAGGAGCT 740  
DB 815 TAGTTCCTCAATATTTTCTCTTGGCAGATCTATGCAATGTCAGATTAACGTAAGATCTCG 874  
QY 741 TTGCGCCCTGCAAAAGGCGCAGATGCGCATCTCGCTCAAGCGGCTACTATGAGCCT 800  
DB 875 ATGAGCGTGTCTGAAAGGGATAGGTGGAATAGTCTGCACTTCACTGATAGAGGCTTC 934  
QY 801 GGGACAGCAATGAGCTCGGAGCAAGAGGCTGCTGAGGAGCGATGGAATTTCACTATT 860  
DB 935 T--TACAACTGACCTGATGACCAAGCAGAGCGCCAAAGAGCGAGGACTTCCATTTG 991  
QY 861 GCTGGTTTGCATATCCCATCTTCTTGAAGAGGACTTATCGAGAGCATGAGAGACGAC 920  
DB 992 GCTGGTTTGTGATC---ATTGATAAGGAGCACTATCCACAGATTAATGCAAGATCTCG 1048  
QY 921 TGGCGAGAGGCTTTCAGCCCTCACTCCCGGAGCTTGGCATCTTCATATGCCGAGAGA 980  
DB 1049 TGAAGAGAGGCTGCCAGGTTCACTCTGAGAGGCTTAATCTGTGAAGGCTCGGACG 1108  
QY 981 CCGACTTCTACGAGTAATTAAC---TACATATCCAGTTGCGCGCACTTGAAGCGTTC 1037  
DB 1109 ACTACATCGGTATTAAGAGTACATCCAGCTTACATGAAAGGCGCAAGACTGATCAGC 1168  
QY 1038 CCGTCCCGGAGACGCAATCTGCGCGCATTCATGACACGAGAGAAATTAAGACGCGCA 1097  
DB 1169 TGGGCGCCAGTACGCTCTGCGGATGCGAGGTTCAATATGTTTTCACGCAATGCGCA 1228  
QY 1098 GCCCGCTTGGGAGAGAGCGCGCTCGCTGCGCTGCGCTGCGCGAGATGTTCCGGA 1157  
DB 1229 AACCATTTGACCAACAGGCGAATTTAATGAGCTTCAATGCGCCGACGCGGATGTAAG 1288  
QY 1158 AGCATCTGCGCGGCTGATGCGCGCTGTAACGCGCAAGC---CCATTTACATCAACGAGAAC 1214  
DB 1289 GGTGCTGATACCTCAAGAGAAATATGGAATTCACAGATCTTACATTAACGAGAACG 1348  
QY 1215 GATGCGCGTCTGAGAGAGAGAAATGATGCTGCGAGAGCGCGTCAAGAGACCCCTTCC 1274  
DB 1349 GAATGACCAAGCTTGA-----AATGACCGGAGACCAATCACTTGGCGACGCGCACGA 1402  
QY 1275 GCATCGGTAATTTACCTGACCTTGACTGATTTCCAGAGGCAATTAACCGAGAGCGCG 1334  
DB 1403 GGTGCGGTTTACAGAGACTACATCGGCGAGCTGAAAGAGGCCATTAACAGCAG---GGAG 1459





; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 33175  
 ; LENGTH: 1932  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: UC-ZMFLMO17071E02\_FLI  
 US-10-425-114-33175

Query Match 11.4%; Score 227.4; DB 12; Length 1932;  
 Best Local Similarity 52.2%; Pred. No. 2e-56; Indels 36; Gaps 9;  
 Matches 735; Conservative 1; Mismatches 637;

QY 81 TGCCCAAGGACCTTGAATGAGGCTTCGCAACGCGCCCTACCATCGATCGAAGCGCGCTCA 140  
 DB 237 TCCCAAGGGGTTCTGTTGCGGAGCGAGCGTCCGATCGAGGCGCGCGT 296  
 QY 141 AAGAGGTGGCCGCGCGCGCTTCATCTGGACAGCTACTGCCACTGGAGCCATCGCGCA 200  
 DB 297 CCACCAAGGCGCGCGCGCGCTTCATCTGGAGTCTTGGCGGACGCTCCAGGAATATATG 356  
 QY 201 CCACAGGCGCGCGCGCGATGCGCTTGCATCACTACCAACCGCTACGATGAGACTTGG 260  
 DB 357 CAGGGAATCAAAATGAGAGAGCTTGCAGATGATCAATACATGCTACAAAGAGAGCGTGC 416  
 QY 261 ATCTCTTGAACCAAGTACGCGCGCAAGGCGCTTCTCTTGTGCTGCTGCGCGATCA 320  
 DB 417 ATCTCATGAAAGTTTGAATCTTGAATGCTTGAATGCTTCAATCTCAATGCTTCAAGATCT 476  
 QY 321 TTCCCTCGGCGCGAGCTGATCCCTCAACGAGAGGGAATGAGTTTACAGCAAC 380  
 DB 477 TCCCGATGCGCAAGG-----AAGTCAATCCAGAGGTGATGCGTATTCATTAATT 530  
 QY 381 TGATTCAGCGCGCTTGAAGCGCGGATATCACGCTTGGGTGACTTGTATCAACTGGATC 440  
 DB 531 TGATTAACCTATCTGCTTGAAGAGGATGATGCTTCAATCAACCTTTATCACTATGATC 590  
 QY 441 TGCCTCAGCGCTTCAAGTCCGCTATGAGGCTGCTCAAGTGGAGAGGTCCAGCTGG 500  
 DB 591 TTCTCTTGGCGCTTGAAGAAATATGAGGCGTGTAAAGCGCAAGATGGCGGACTTG- 649  
 QY 501 ACTTCAGCGGATGAGAGGTTGCTTGAAGCTTTGGGAGCGAGTCCAGAACTGGA 560  
 DB 650 --TTTACAGACTATGCTGACTTGTGTTAAAGCTTACGCGCATGCGGTAAAGCACTGGT 707  
 QY 561 TCACCATCAACGAGCCCTGATTCAGGCACTATGATATGCAACGCGAGCAAGCGCC 620  
 DB 708 TTACATTCATGAGCGCAAGATATGATGCGTACTTGGCTATGACAGAGGTCAAAATCTTC 767  
 QY 621 CGGCGAGAGCGACATTAACAGCACTCCACGAGGCGCAACATGCGCACTGAGCGGTGGC 680  
 DB 768 CTCGAAGGTGC-----ACCAGATGCGCTGCTGTGGGAATTCAGAACCGAACCCTTACA 821  
 QY 681 TCGCTGGAAGAGCGCGAGATCATGAGCATGCGCGCGCGCTGCTCAAGAGGAGCT 740  
 DB 822 TAGTTGCTCAATATTTTCTTTGGCACATGCTACTGCAATGGAAGATACCGATCGAAT 881  
 QY 741 TTGCGCCCTCGCAAAAGGCGCAATGCGCATCTCGCTCAACGCGCATCTATGAGCCCT 800  
 DB 882 ATCAGGCTGCTCAAGAGGTAAAGGTGGAATGATGCTGAGCTTCAACTGGTACGAGGCTC 941  
 QY 801 GGAACGACATGAGCTTCGCGCAAGAGAGCTGTGACGACGAGATGGAATTTCACTTG 860  
 DB 942 T--TACAACTCACTGATGACCAAGCAGCGCCAAAGAGCGAGGACTTCCACTTG 998

QY 861 GCTGTTTGCATATCCCATCTTCTGAGAGAGACTATCCAGAGACATGAAAGACAGC 920  
 DB 999 GCTGTTTGTGATTC--ATTGATTAAGAGACATATCAAGATTAATGCAAGATCTCG 1055  
 QY 921 TGGCGAGAGGCTTCAGGCTTCACTCCCGGAGCTTTGCCATCTCAATGCGCGAGAGA 980  
 DB 1056 TGAAGAGAGGCTGCGCAGGTTCACTCTGACAGGCTTAACTGTGAAAGGCTCGGCG 1115  
 QY 981 CCGATCTTACAGGCAATGATTAAC--TACACATCCAGTTGCGGCGCCACTAGACGCTC 1037  
 DB 1116 ACTACATCGGTATCAACAGATACATCACTACATGAAGGGGAGAGAGCTGTCCAGC 1175  
 QY 1038 CCGTCCCGGAGAGGAGCTATCTCGGCGCATTCATGAGCAACAGAGAAATGAAGACGGA 1097  
 DB 1176 TGGCGCGAGTACTCTGCTGCGGATTTGGCAGGTTCAATATGTTTGGACGCAATGGA 1235  
 QY 1098 GCGCCGTTGGCGAGAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157  
 DB 1236 AACCGATTGAGACCAAGCGGAAATTTAAGTGGCTTACATGCGCCGAGCGGATATAG 1295  
 QY 1158 AGCATCTCGCGCGGATGACGCGCTGACGCGCAAGC--CAATCTATATCACCGAAGAG 1214  
 DB 1296 GGTGCGGAACTACCTCAAGAGAGATGATGAGATCAAGATCTTACATTAAGGAGAGAG 1355  
 QY 1215 GATGCGCGCTCGGAGAGAGACATGACGTGCGAGAGGCGCTCAACGACCTCTTC 1274  
 DB 1356 GAATGAGACGAGCTTGA-----ACCTTGACCGAGACAGTACCTGCGGAGCGCACGA 1409  
 QY 1275 GCATCGGTACTTTGACTCGCACTTGGACCTGCAATTTTCAAGGCGCATTAACGAGAGCG 1334  
 DB 1410 GGGTGGGTTCTACAGAGAGTACATGCGCGAGCTGAAGAGGCGCATGAGACAG--GAG 1466  
 QY 1335 TCGCTCAAGAGGATCTTTCGTGGCGCTTGTCTGATTAATTGAAATGTCAGATGCT 1394  
 DB 1467 CGAAGCTGCTGCTGATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1526  
 QY 1395 ACGAGCCGAGATTCGCGCTGACGTTCAAGACTACACCTACCTCAAGCGCGCAAGA 1454  
 DB 1527 ACTGCTCAAGTTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1586  
 QY 1455 AGTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1483  
 DB 1587 GGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1615

RESULT 9  
 US-10-425-114-25892  
 ; Sequence 25892, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 25892  
 ; LENGTH: 1737  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LTB3960-023-C7\_FLI  
 US-10-425-114-25892

Query Match 11.0%; Score 219.2; DB 12; Length 1737;  
 Best Local Similarity 51.3%; Pred. No. 5e-54;

Matches	727; Conservative	1; Mismatches	654; Indels	36; Gaps	8;
Qy	81	TGCCAAGCATTGATGGGCTTGCACAGCGCGCTACCGATGAGAGGCGCTCA	140		
Db	228	TTCCCAAGGGGCTTGTGTTGGGAGCGGCGGTGGCTTACAGGTGAGGAGCATGGCGC	287		
Qy	141	AAGAAGTGGCCCGGCGCCGCTTCATCTGGGACGTAATCCCACTGGAGCCATCGGCA	200		
Db	288	ACAAGAGCGCGCGCGGCGGAGCATTTGGGACGCTTATCAATGATCCCGGGGAGATCG	347		
Qy	201	CCAAAGGCGGCAAGCGGATGTGGCTTGCATCAACACCGTACGATGAGAGCTTTG	260		
Db	348	CAAAACAAGCGCGCGGAGCTGTGACGATGATCCCTACAAAGAGAGACCTGA	407		
Qy	261	ATCTCTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTGTGTGTGGTCCGATCA	320		
Db	408	ATATCATGAAATATATGGGTTGACGCTTACCGGTTTGTGATCTTGTGTCAAGATAT	467		
Qy	321	TTCCCTCGGCGGAGGCTGGATCCGTCACAGAGAGGAGATTTGATTTTACAGCAAC	380		
Db	468	TCCCAATGGAAC-----GAGAGTGAATGGAAGAGAGTGGGTACTACACAGCGC	521		
Qy	381	TGATTTAGCCCTTGTGAGCGGGGATACAGCTTGGGATCTTGTACCATCGGATC	440		
Db	522	TGATTAACATCATGTGTAAGAAAGGATCAACCTTACGCAACCTGTACCATGACCC	581		
Qy	441	TGCTCAGGCGCTTACGATCGCTATGAGAGGCTGGCTCAAGCTGGAAGAGTCCAGCTG	500		
Db	582	TACCGAGGGGCTGAGAGTCCGCTACGAGAGGCTGTGACAGAGAGTGT--GAGAT	638		
Qy	501	ACTTTGAGCGGATGAGAGGTTGTGCTTTGAAAGCTTTTGGGAGCCGAGTCCAGAACTGA	560		
Db	639	CGTTGCAGACTACGCGACTTCTGTCTGGGGCGTTCCGCGAGAGGGTAAABAATGGC	698		
Qy	561	TCAACATCAACGAGCCCTGATTCAGGCCATCTATGATATGCCACCGGAGCAAGCC	620		
Db	699	TGACGTTCAACGAGCGCGGGGTGTGGCCCTTACGAGGTACGACGAGGAGTTGGCGC	758		
Qy	621	CGGGGAGGAGACATTTAACAGCACTCCACGAGGCGAACATGCACTGACGAGCCGTGGC	680		
Db	759	CGGGAGGTGCAAGGGGTGCGA-----GCCGGGGGGAGCTGGGAGCCGAGCCCTACG	812		
Qy	681	TGCTGGAAGAGCCGAGATCATGAGCATGCCGCGCGCTGGGCGCTTACAGAGGAGCT	740		
Db	813	TGCTGGGAGCAACCTCATCTCTCCACGCGCGCGCTCCAGAGTACCGCGCAGGC	872		
Qy	741	TTGCGCCCTTGGCAAAAGGCGCAGATCGGATCTGCTCAAGCGGCACTATGAGCCCT	800		
Db	873	ACCAGCGCAGCGAGAGGGGAGGGGTGGGATCTGCTGAGATTTGATGTGTAAGAGCCC	932		
Qy	801	GGGACAGCAATGAGCTCGGAGACAGAGGCTGTGAGCGACGATGGAATTTCACTTG	860		
Db	933	TCAAGCGGAGCTCAGCGCGCGAGCGCGCGCTCAMAAGTCCAGAGATTTCCAGCTCG	992		
Qy	861	GCTGTTTCCCATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAAGACAGC	920		
Db	993	GATGTTCTGTGACCCCATCTCT---ACGGGAGTACCCCAAGTGGGTCCGAGAAAGC	1049		
Qy	921	TGGGCGAGAGGCTTCAAGCCCTCACTCCGCGGACTTTGCTCATCTCAATGCCGAGAG	980		
Db	1050	TCAAGGCGAGGCTCCCAAGTTCAAGGCTGAGGA---GGCGGCTTCAATGCCGAGCTCCA	1106		
Qy	981	CGGACTTTCAAGGAGATTAATTAACAATCCAGATTGCGGCGCCACTTAAGAGAGTCCG	1040		
Db	1107	TGCACTACGTGGAGTCAACAGTACATGCTTCACTACGTCGTGATGAGCGCCAAAGC	1166		
Qy	1041	TCCCGAGAGC-----GACTATCTGGCGCCATTCATGAGACACAGAGAAATAGG	1091		
Db	1167	CTACGGCGGGCGCGCCAGCTACTCGTCCGACTGGCAGCTGAGTGTGTATGAAGCGC	1226		
Qy	1092	ACGGAGAGCCCGTTGGCGAGAGAGCGGCTGCTGGCTGCGCTCTGCGGAGCATGT	1151		
Db	1227	ACGGTGTGCGATTTGAGCAAGAGGCGAATCAAGTGGCTTCACTGTGCTTTGGGAGC	1286		

Qy	1152	TCCGAAGACATCTGCGCCGGGTGTACGCGCTGTACCGCAAGCCCATCTACACGAGA	1211
Db	1287	TGTACAAAGCCGTACACTTACATCAAGAGAGATGACGCAACCCCA---CGATGCTCTGT	1343
Qy	1212	ACGAGTCCCGTGGCCCTTGAGAGAGAACATGACGTGCGAGAGGCGCTCAAGCCCT	1271
Db	1344	CAGAGAACGGTATGACGAGACCCGCGGCAAGTCAAGTGGGCGAGGCGGTGCGGAGCGG	1403
Qy	1272	TCCGATCCGCTACTTTGACTGCACTTGGACTGCAATTTCCAAAGGCAATTAACGAGAGC	1331
Db	1404	CGAGGTGGCTTACT---ACCGAGCTTACGTGCGCAGCTGAAGGCGCGCATGAGCGCG	1460
Qy	1332	GGCTGCTGTCAGAGGGGTACTTTGCGTGGGCGCTTGTCTGATTAATTTGAAATGCTAGATG	1391
Db	1461	GGCGCAACGCTGCGGAGTACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1520
Qy	1392	GCTACGAGACCCAGATTCGCGGCTGACGTTCAACAGACTTACACACCTCAAGCGAGCCCA	1451
Db	1521	GGTACACGTCCTCGGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1580
Qy	1452	AGAGTCTGCGCTGCTGCTTCAAGAGACATGTTGGCGC	1489
Db	1581	AGAGCTCGGCTGCTGCTGCTTCAAGAGATGATGATGCGCGC	1618

RESULT 10  
US-10-425-114-13710  
; Sequence 13710, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaka, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT FILING DATE: US/10/425,114  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 13710  
; LENGTH: 1759  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB143-030-G7\_FLI  
US-10-425-114-13710

Qy	199	CACCAAGGCGCAAGGAGTGGCTTGCATCACTACACCGCTAGAGTGAAGACTT	258
Db	368	CGCAACCAAGCGCACCGGAGCTGACTTTGACGATCACTGCTACAGAGAGAGCT	427
Qy	259	TGATCTTTGACCAATGAGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGGAT	318
Db	428	GATATCATGAAATATATGAGGTTGACGCTTACCGGTTTGTGATCTTGTGCAAGAT	487
Qy	319	CATTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGAGATTTTACGCA	378
Db	488	ATTCCTCAATTTGAAACCGAGAA-----GTGAATGGAAGAGTGGCTTACTACAAAG	541
Qy	379	ACTGATTGAGCGCTGTGAGGCGGGTATCAAGCTTGGGATGACTTTGACACTGGGA	438
Db	542	GCTGATTAATCTATGTGAAGAAAGGATCAACCTTACGCAACCTGACACTACGA	601
Qy	439	TCTGCTCAGCGGCTTCAAGATGCTATGAGGCTGGCTCAAGTGAAGAGTCCAGCT	498

```

Db      602 CTTACCGGAGCGCTGAGAGTCCGGTACGAGGCGCTGTTGACGAGAAATGCT---GAG 658
Qy      499 GGACTTTGAGCGGATATGAGAGTTGTGCTTTGAAAGTTTGGGAGCCGAGTCCGAACCTG 558
Db      659 ATCGTTGCGAGACTACCCGCACTTCTGCTTGGGGCGCTTCGGCGACAGGGGGAAGACTG 718
Qy      559 GATACACATCAAGAMCCCTGATTTCAAGGCGCATATGATGATATGACCCGCGACCAAGCG 618
Db      719 GCTGACCTTCAACAGACCGCGGTTGTGGCCGCTTGAAGTACACAGCGGACAGTTCCG 778
Qy      619 CCCGGGAGAGAGCACTTACCAAGCACTCCACGAGGCAAACTGCTGACTGAGCCGTG 678
Db      779 GCGGGGAGGTGCACGGGGTCGAGG-----CCGGGGGGGAGCTCGGGCACGAGCCCTTA 832
Qy      679 GCTGCTGGAAGAGCCCATATGAGCCATGCCCCGCGCGCTGAGCCGTTACACAGGGA 738
Db      833 GGTGTTGGCGGACCACTCATCTCTCCACGCGCGCGCTCCAGAGTACCCCGGAG 892
Qy      739 CTTTGGCCCTTCGCAAAAAGGCGCAGATCGGCACTCGCTCAACGCGCACTACTATGAGCC 798
Db      893 GCACACAGCCACGAGAGGGGCAAGGTCGAGATCTGTGATTTCTGTGTGTAAGACC 952
Qy      799 CTGGGACAGCAATGAGCTCGGACCAAGAGGCTGCTGAGCGACGAGATGAAATTCACAT 858
Db      953 CCTCACGGCGGACTCAGCGCGCGACCGGGCGCGCTCAAGGTCCAGAGACTTCCAGCT 1012
Qy      859 TGGCTGTTTCCCAATCTCTTGTGAAGAAGACTATCCAGAGACTGAAAGAAACA 918
Db      1013 CGGATGTTCTTGCACCCCATGCTCT---ACGGGAGATACCCCAAGTGGTCCGAGAG 1069
Qy      919 GCTGGGAGAGAGGCTTCAGCCCTCACTCCGCGGACTTTGCACTCCATGCGCGAGGA 978
Db      1070 GGTCAAGGAGCAGGCTCCCAAGTTTCAAGGCTGAGGA---GGCGGTCTATGTCGAGGCTC 1126
Qy      979 GACCGACTTCTACGCGCATGAAATTACTACACATCCCAATTCGCGCGCACTTACAGAGTCC 1038
Db      1127 CATGCACTACGTCGAGAGTCAACAGTACACTGCTACTAGCTGCTGATTCAGCGCCAAA 1186
Qy      1039 GGTCCCGAGAGC-----GACTATCTGGGCGCATCCATGAGCACAGAGAAATA 1089
Db      1187 CGTACGCGCGCGCGCCCACTCTCTGCTGACTGCGACGCTGATGTTGTTAAGACG 1246
Qy      1090 GGAAGCGAGCCCGCTGCGAGAGAGAGGCGCTCGCTGCGTGGCTCTCCCGCGACAT 1149
Db      1247 CGAGGTTGTGCGATTTGACCAAGGCGCAACTCAGACTGTGCTTACATCTGTGGGG 1306
Qy      1150 GTTCCGGAAGCATCTCGCCGAGGTGTACGACCTGTACGCGCAAGCCCATCTACATCAGCA 1209
Db      1307 ACTGTACAAAGCCGCTCACTAGTCAAGAGAAAGTACGCGCAACCCACAGATGCTCTGTGTC 1366
Qy      1210 GAACGGAATGCCCTGCTGAGAGAGAAACATGACGTGCGAGAGGCGCGTCAAGCAACC 1269
Db      1367 AGAGAACGATATGAC---GACCCGCGGCAAGTCAAGTGGGCGCGGCGTGGCGAGACG 1423
Qy      1270 CTTCCGCACTCGGTTACTTGAATCGCACTTGGAGTCAATTTCAAGGCGCATTTACCAAGGA 1329
Db      1424 GCGGAGGTTGCGCTACT---ACCGAGACTAGTCCCGCAGCTGAAGGCGGAGTGAAGCG 1480
Qy      1330 GCGCGTCTGTCTCAAGGGGTAATTTGCTGTGGCGTGTCTGTATTAATCTTGAATGTCTCA 1389
Db      1481 GCGGCGCAACTGCGTGGATATCTTGTCTGTGCTGTGCTGACAACTTCGATGAGTGAAGCT 1540
Qy      1390 TGGCTACGAGACCCAGATTCGCGGTCAAGTTCACAACTACACACCCCTCAAGCGGACGCC 1449
Db      1541 GGGGTACAGTTCGCGTGTGCGCTGTCTACGTCACTTCAGAGAGGCTCCGCGCTACCC 1600
Qy      1450 CAAGAAGTCTGCCCTGTCTCTCAAGACATGTTTCCGCC 1489
Db      1601 CAAGAGCTCGGCTACTGTTCAAGGATGTATCCCGGCG 1640

```

RESULT 11  
US-10-425-114-24850

```

: Sequence 24850, Application US/10425114
: Publication No. US20040034888A1
: GENERAL INFORMATION:
: APPLICANT: Liu, Jindong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovalic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: FILE REFERENCE: 38-21(5313)B
: CURRENT APPLICATION NUMBER: US/10/425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ. ID NOS: 73128
: SEQ ID NO 24850
: LENGTH: 1735
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3689-207-B2_FLI
US-10-425-114-24850

```

Query Match 9.3%; Score 184.6; DB 12; Length 1735;  
Best Local Similarity 49.8%; Pred. No. 9.2e-44;  
Matches 733; Conservative 0; Mismatches 704; Indels 36; Gaps 9;

```

Qy      77 GCTTGCCCAACGACTTTGAATGAGGCTTTCGCAACGCGCCGCTTACAGATGAAAGCGCC 136
Db      166 GCGTTCCCGAGCGGCTTGTCTTGGAGCGCGCGGTGGGTACAGTGCAGAGGGATG 225
Qy      137 GTCAAGAGAGTGGCGCGCGCGCTCCATCTGGGACACGTAAGTCCACTGAGCCATTCG 196
Db      226 GCAAGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 285
Qy      197 CGACCAAGCGCGCGCAAGCGGATGTGGCTTGGATCACTACCAACCGCTACGATGAGAC 256
Db      286 ATCCCTAACATGTGCGACCGCTGACGTGACGTCGACAGCATATCTCCGTACAAAGAAAT 345
Qy      257 TTTGATCTTTGACCAAGTACGCGCGCAAGGCGCTTCTGCTGTGTGTCGCGG 316
Db      346 GTGAACATATGAAGAAACATGGCTTTGATGTCGTAACGAGTTTGATCTCTTGGTCAGG 405
Qy      317 ATCATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376
Db      406 ATTTCACAGATGGAAGTGGGAG-----GTGAACAGAGAGAGTGAATTAATTAACAC 459
Qy      377 AAACGTATGACGCTCTGTGAGCGGGGTATACGCTTGGTGAATTTGTATCACTGCG 436
Db      460 AGGCTCATATGATTAATGCTTCACAGCAAGTATCCGCGGTATGCAAAATCTTACCATAT 519
Qy      437 GATCTGCTCAGGCGCTTTCAGATCGCATGAGAGGCTGGCTCAAGTGAAGAGTCCAG 496
Db      520 GACTCCCATTTGGGACCTTCATGAAGACAGTCTGGGCTGGCTTAC---CCAAAGATTGTG 576
Qy      497 CTGACTTTGAGCGGTAATGCGAGTGTGCTTTGAAGCTTTGGAGCCGAGTCCAGAAC 556
Db      577 GAGGCGTTGAGACTAGTACGCGGAGTTCTGCTTCAACGCGTTCCGAGACAGGTTGAAGAC 636
Qy      557 TGAATACCATCAAGAGMCCCTGATTCAGGCCATCTATGATATGCAACCGGACAGAAC 616
Db      637 TGTATTACTTTCACAGAGCCAGGTGCGTCTCTCTGCGCTACGAATGAGCTTTCGAC 696
Qy      617 GCGCGCGCGGAGACAGATTAACAAGACTCCACGAGGCAACACTGCGACTGAGCGG 676
Db      697 GCACCGGAGAGGT-----TTCGGGTGCGCGCGCGGAGCACTCACCACGAGGCGG 750
Qy      677 TGGCTCGTGAAGAGCCAGATCATGAGCCATGCGCGCGCGCGCGCGCTTACAGCAGG 736
Db      751 TACCTTGTGCAACATCTCATCTTCTATCAGAGTGTGCGTCAAGCGATTCAGCGAC 810
Qy      737 GACTTTGCGCCCTGCAAAAGGCGCAAGTGGCATCTTGCTCAACGCGGACTACTATGAG 796

```

Db	811	AAGATATCAGCTTCACCAAGAAAGGGGAAGATGGAATTTCTCTGGATTTTCGTGTGATACGA	870
Qy	797	CCCTGGGACAGCAATAGAGCTTCGGGACAAGAGGCTGTGACGACGATGGAATTTCA	856
Db	871	CCTT---TCAGGACAGCAATGCGAGACGAGCTGCAGCACACGAGGCCAGGAACTTCCAC	927
Qy	857	ATTGGCTGTGTTGGCCAAATCCCATCTTTCTTGGAAGAAAGCATATCCAGAGCATGGAAGA	916
Db	928	CTAGGCTGTGTTCTTTGACCCCA---TTGTACATGAGCGGTACCCGTAATCTGATGCAAGAG	984
Qy	917	CAGCTGGGCGAGAGGCTTCAGACCTCCACTCCCCGGGACCTTTGGCATCTCTCATGCGGGA	976
Db	985	ATTGCCAAAGACAGGCTACCGTTGTTCACGATGGAAGAACAGAGATGGTGAAGGCTCT	1044
Qy	977	GAGACCGACATTCTACGGGATGAATTAACAATCCAGTTGCGCGCCACCTTAGACGGT	1036
Db	1045	ATTAGACTATGTGGCATCAACACATACTTCTTCTCATGTAAGGACCTGGGACATGG	1104
Qy	1037	CCCGTCCCCGAGACGGAATATTTGCGCGCATTCATGAGACCAAGAAATA---GGAC	1093
Db	1105	AACCTGAGCGCAGTACGAGCTTACAGGATGATGGCATGTTGGTTGTCTACGAAAGCAAT	1164
Qy	1094	GGGACGCCCGTTGGGAGGAGAGCGGCGCTCGCTGGCTGCGCTCTCGCCGACATGTTTC	1153
Db	1165	GGAGTCTTATTGGGCGCTCACGCAAACTTCTACTGGCTTACATTTGTGCTGGGGCATTC	1224
Qy	1154	CGGAAGCATCTCGCCCGGGGTATACGGCGCTGTACCGCAAGCCCATCTTACATCA---CCGAG	1210
Db	1225	AACAAGCTGTACGCTATGTACAAAGAACTTACAATAATCTTACATGATCTTGTCTGAA	1284
Qy	1211	AACGATGCCCGTCCCTGGGAGAGGAAACATGACGTGCGAGAGAGCGCTCAACGACCCC	1270
Db	1285	AACGAAATGAGCAACACTGTGTATGTACATATTACT-----CAGGGTGTGATGACACA	1338
Qy	1271	TTCCGCAATCCGCTTCTTGTGATCTCGACCTTGACATGATTTCCAAAGGCCATTAACCGAGAC	1330
Db	1339	GTAAGATCTCGTTATTACAGAGACTACATACATGAGCTCAAGAAAGCAATA---GATGAT	1395
Qy	1331	GCGCTGTCTGTCMAAGGGGTACTTTTGCCTGGGCGCTTCTCGATTAATTGGAATGTCAGAT	1390
Db	1396	GGTGCACAGATCATTTGGGTACTTTGGTGTGTGCGCTTGACACATTTGAGTGGAGGCTT	1455
Qy	1391	GGCTACGAGACCAAGATTGGCGCTCAAGCTTCAAGACTACACACCTTCAAGCGCACGCC	1450
Db	1456	GGGTACACTTGGCGGTTTGGCTTGGTGTACGTGGACTACAGAGCTTGAAGAGGTACCCC	1515
Qy	1451	AAGAAATGTGCGCTGCTCTCTCAAGACATGTTTGGCGCCCGGACAGAGGCTTAAGTGGG	1510
Db	1516	AAGGACTAGCTTTTGTGTTCAAGCATATGCTGTCCAAAGAAAGGCTTAGAATTTGCA	1575
Qy	1511	GCATTAAGAAAGGAAATTTCTTCTTGACATCA	1543
Db	1576	GACAAAGAGGACCACTGGCTTCAAGTGTCTATCA	1608
RESULT 12			
US-10-425-114-27741			
Sequence 27741, Application US/10425114			
Publication No. US20040034888A1			
GENERAL INFORMATION:			
APPLICANT: Liu, Jingtong			
APPLICANT: Zhou, Yihua			
APPLICANT: Kovalic, David K.			
APPLICANT: Screen, Steven E			
APPLICANT: Tabaska, Jack B			
APPLICANT: Cao, Yongwei			
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With			
FILE REFERENCE: 38-21(53313)B			
CURRENT APPLICATION NUMBER: US/10/425,114			
CURRENT FILING DATE: 2003-04-28			
NUMBER OF SEQ ID NOS: 73128			
SEQ ID NO 27741			

Query Match	9.3%	Score 184.6	DB 12	Length 1757
Best Local Similarity	49.8%	Pred. No. 9.2e-44		
Matches 733	Conservative	0	Mismatches 704	Indels 36; Gaps 9
US-10-425-114-27741				
Query	77	GCTGTGCCCCAACGACCTTTGATGGGGGCTTGCGAACGGCCGCGCTCCACGATCGAAGCGCC	136	
Db	188	GCGTTCGCCGACGGCTTCCTTCCTTGGGACGGCCGCGCTCCGCGTACCGAGTGCAGGGCATG	247	
Qy	137	GTCMAAGAAGGTGGCCGCGCCCGCTCCATCTGGGACACGTAATGCCACCTTGAGCCATCG	196	
Db	248	GCCAAAGACGGCGGGGGGGCCCGACGATCTGGGACGCGCTTCATAGAGGTTCTTGAGACC	307	
Qy	197	CGCACCAAGCGCCCAACGGCGATGTGGCTTGGCATCACTACACCGCTACGATGAGAC	256	
Db	308	ATCCCTTAACAATGCGACCGCTGAGGTACGCGTACGACAGTATCATCTCGTACAAAGAAAT	367	
Qy	257	TTTGATCTCTTGACCAAGTACGGCGCAAGGCCACCGCTTCCTCTTGCTGGTGGCGG	316	
Db	368	GTCACATTAATGAAGAACATGGGCTTTGATGCGTACCGGTTTGGATCTCTTGATCGAGG	427	
Qy	317	ATCATTCGCCCTCGCGGCGAGCTGATCCGCTCAACGAGAGGGAATTGACTTTTACAGC	376	
Db	428	ATTTTCCAGATGGAACCTGGCAG-----GTGAACAGAGAGAGTGGATTACTACAC	481	
Qy	377	AAACTGATTAAGCGCCCTGTTGAGGCGGGGATATACGCGCTTGGTGACTTTGTACACTGG	436	
Db	482	AGGCTCATAGATTATCATGCTCCACAGAGTATCCGCGGTATGCAAAATCTCTACATTAT	541	
Qy	437	GATGTGCTCAAGGGGCTTTCACGATCGCTATGGAGGCTGGGCTCAACGTGAAGAGGTCCAG	496	
Db	542	GACCTCCCATTTGGACATCTCCATGAACAGTACTGGGCTGGCTTAC---CCAAAGTTGTG	598	
Qy	497	CTGCACTTTGACCGGTATGCGAGGTTTGCTTTGAAAGTTTTGGGACCGAGTCCAGAAC	556	
Db	599	GAGGCGTTTGGACACTACGCGAGTTCTGCTTCCACCGCTTCCGAGACAGGGTAAAGAC	658	
Qy	557	TGATATCAACATCAACGACCCTGGATTCAAGGCATCTATGGATATGCCACCAGCAGAAC	616	
Db	659	TGTTTACCTTCAACGAGCGAGGTGTGCTGCTCTGGGCTACGACAAATGAGCTTGAC	718	
Qy	617	GCCCCGGGACGAGACGATTAACAAACACTCCACGAGGCAACACTGCCACTGAGCGG	676	
Db	719	GCACCGGAGAGT-----TTCGGGTGCCCCCGGAGGCAACTCACACGAGCGG	772	
Qy	677	TGGCTGCTGGAAAGGCCAGATCATGAGCATCCGCGCGCTGGGCGCTTACAGCAGG	736	
Db	773	TACCTTGTGCGACACATCTCATCTCTTCTCATGAGCTGGGGTCAAGGCATACCGCAC	832	
Qy	737	GACTTTGCGCCCTGCAAAAGGGCCAGATGGCATCTCGCTCAACGGCGACTACTATGAG	796	
Db	833	AAGTATCAGCTTCAACGAAGGGGAGATTGGAAATTCCTCGGATTTCTGTGTGTAAGAA	892	
Qy	797	CCCTGGGACAGCATGAGCCTCGGGACAAGAGGCTCTGAGCGACGAGATGGAATTTTAC	856	
Db	893	CCCT---TCACGACAGCAATGCGAGACGAGCTACACACAGCGACGAGGACCTTCCAC	949	
Qy	857	ATTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTTACAGAGAGCTAAGAGAG	916	
Db	950	CTAAGGCTGTTCTTGAACCCA---TGTATCATGAGACGATACCGGTACTGATTCGAAGAG	1006	
Qy	917	CAGCTGGCGAGAGCTTCCAGCCCTCACTCCCGGAACTTTGCAATCTCAATGCGGGA	976	
Db	1007	ATTGCAAAAGACAGGCTACCGTTGTTTCAGCGATGAAGAGCCAGATGTGTAAGAGCTCT	1066	
Qy	977	GAGACCGACTTTCAGGCGATGAATTACTACATCCAGTTCCGGCGCCCACTTAGACGT	1036	



QY 1274 CGCATCCGCTACTTGTGACCTGCACTTGCATTTCCAAGGCCATTATCCAGAGCGC 1333  
 Db 1144 CGGATGCCCTACCTGACCTGCACTGCGGCGCGGACCGCTGCATT---GAGCGCGC 1200  
 QY 1334 GTGCTCTCAAGGGGTACTTGTGCGGGCTTGTCTGATAAATTGGAATGTCAGATGCG 1393  
 Db 1201 GTGCGCTCAAGGGGCTACTTGTGCGGGCTTGTCTGATAAATTGGAATGTCAGATGCG 1260  
 QY 1394 TAGGAGCCAGATTTGGCGGTCAGCTTCAAGACTACACCACTCCAGAGCGGACGCCAAG 1453  
 Db 1261 TAGGAGAGCGGTTGCGGCTGTCAGCTGAGACTACGAGAGCGAGAGCGGCGGAG 1320  
 QY 1454 AA 1455  
 Db 1321 GA 1322

RESULT 14  
 US-09-938-842A-1282  
 ; Sequence 1282, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Kreps, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SAME, AND METHODS OF USE  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938, 842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 1282  
 ; LENGTH: 1734  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-938-842A-1282

Query Match 9.28; Score 182.8; DB 9; Length 1734;  
 Best Local Similarity 49.28; Pred. No. 3.1e-43;  
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

QY 79 TCTGCCAAGCATTTGATGATGGGCTTCGCAAGCGCGCCCTACAGATCGAAGGCGCGT 138  
 Db 90 TTTTCTGATGATTTTCTTTTGGGACAGCTGCTCGGCGTTTCAGTACGAAGGTGCAC 149  
 QY 139 CAAGAAGGTGCGCGCGCCGCTTCATCTGGACACGTAAGTCCAGCTGAGCGCATCGC 198  
 Db 150 AAGTGAAGGTGGCAAGTCTCAACTATATGGGATCACTTCAAGCTTCAGTATCCAGAAG 209  
 QY 199 CACCAAGGCGCCCAAGCGGATGCTTGGGATCATACCAAGCTTACGATGAGACTT 258  
 Db 210 GACCAAAATGATATGCAATGATGAGCAATGATTTTATCATCTGTACAAAGATGCAT 269  
 QY 259 TGATCTCTTACCAAGTACGCGCAAGCGCTACCGCTTCCTGCTGCTGCTGCGGAT 318  
 Db 270 AAAATTGATAGAGAGCTAAACATGACGCTTTCGATTTTCAATCTCTGCTGCTGAGATT 329  
 QY 319 CATTCCTTGGCGGCGGAGGCTGATCCGTCACAGAGAGGGAATTGATTTTACAGCAA 378  
 Db 330 AATAACCAAGTGAAGAGTAAAGATGAGTAAACAAAGAGGTGTACAAATTCTACAGGA 389  
 QY 379 ACTGATGAGCGCTGTTGAGGCGGAGTATCAGCTTGGGATCTTGTATCACTGGA 438  
 Db 390 TCTCATAGAGCAACTTCTGCTAATGACATCAACCTTGCATGACGCTTATCATTTGGGA 449

QY 439 TCTGCTCAGGCGCTTACAGATGCTATGAGAGCTGCTCAACGTGAAGAGTCCAGCT 498  
 Db 450 CCAACCAATCTTTGAGAGCAATATGTGCTTTTAAAGCCCTTAAATCTGAGA--- 506  
 QY 499 GGACTTGAAGGATATGCGAGTTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTG 558  
 Db 507 AGATTTTCAGATTTTGAAGAAATTTGTTGAAGGTTTGAAGATTAAGTTAAGTGTG 566  
 QY 559 GATCAGCATCAAGAACCCCTGATTCAGGCTCATCTATGATATGATGCAAGCGGACAGCG 618  
 Db 567 GACAAATCAATCAAGAACTTTATATGACTGTGGGCTTATGATCAAGGTAAACAGGC 626  
 QY 619 CCCGAGCAGAGC-----AGCATTAACAAGCACTCCACGAGGCGCAACTGCACTGA 672  
 Db 627 GCTGGAGCATGCTCAAAATAGGTTAAACGAAAGTGTACAGGCTGAGATTGAGTACCGA 686  
 QY 673 GCGGTGCTGCTGGAAGGCGCAGATCATAGCCATGCCGCGGCTGCTTACAG 732  
 Db 687 GCTTACATTTGTTTCAATCACAATCTTCTTGGCCATGCGCTGAGTGAAGAAATTTG 746  
 QY 733 CAGGACATTTGCGCCCTCGCAAAAGGCGCATGCGCATCTGCTCAAGCGGCACTACTA 792  
 Db 747 AATAATGAAAAAATCTTGCATGATGCGCAAAATGGATGATCTATACCAAGATGTT 806  
 QY 793 TGAAGCTTGGGACAGCAATGAGCTTCGCGACAAAGAGGCTGTAGAGCAAGATGAAAT 852  
 Db 807 CGAGCTTATCATTTCCGATTCACATGACATTAAGAAAGAGCTGAAAGAGCTTTGCTT 866  
 QY 853 TCAATTTGCTGTTTGGCAATCCATCTTCTTGAAGAGAGCTATTCAGAGCATGAA 912  
 Db 867 TGAATTTGATGCACTTGTATCACTGATCACTTC---ACGAGATTAATCCAGAGATTG 923  
 QY 913 GAAGCACTGGGAGAGAGGCTTCAGAGCCCTCACTCCGCGGAGCTTTGGCATCTCAATG 972  
 Db 924 AAGATGCGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983  
 QY 973 CGAGAGACCGAATCTTACGAGCATGATTAATTAATTAATTAATTAATTAATTAATTA 1032  
 Db 984 TTCA---TGAGATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1040  
 QY 1033 CGGTCCGCTCCCGAAGCGAGCTATCTCGGCGCATCATGA-----GACACAGA 1083  
 Db 1041 TCACATAGACCTCGAATAAATCTGTTCAAAATGACACCATGAGATGAAATGAACTGAC 1100  
 QY 1084 GAATTAAGACCGAGCGCCGTTGGCGAGAGAGAGCGGCTCGCTGCTGCTGCTGCCC 1143  
 Db 1101 TAATCAAGTGTGATCATATCTGAGCTGGGAGAAAGGCGCTTCTTATTTTCAACACC 1160  
 QY 1144 GGAATGTTCCGAGACATCTCGCCGAGTGTACGCTGTAC---GCAAGCCCATCTTA 1200  
 Db 1161 GGAAGCTTACGAAAGTTCTTAACTATTAATTAAGAGATTAATTAATTAATTAATTA 1220  
 QY 1201 CATACCGAAGCGATGCGGCTGCTGAGAGAGAAACATGACGTCGAGAGGCGCT 1260  
 Db 1221 CATCAAGAAATATGATTAATGACAAACGACGCTTACAAACCAAGAGAAATTTGT 1280  
 QY 1261 CAACGAGCCCTTCCGATCCGATCTTGAATGCACTGCACTTGAATGCAATTTCCAGGCCAT 1320  
 Db 1281 GAAGGACACATTTAAGATTAATGATTAATGACATTAATGACATTTGGAAGAACTTCA 1340  
 QY 1321 TACCAAGAGCGGCTGCTGCTCAAGGAGTACTTGTGCGGCTTCTGATTAATTTGA 1380  
 Db 1341 AGTGAAGATGGGTGAGATTAAGAGATTAATTAATTAATTAATTAATTAATTAATTA 1400  
 QY 1381 ATGTCAGATGCTACGAGCCAGATTGCG 1410  
 Db 1401 ATGAGAGCATGATTAATGATGCAAGATTGCG 1430

RESULT 15  
 US-09-938-842A-1282  
 ; Sequence 1282, Application US/09938842A  
 ; Publication No. US20040009476A9



```

; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1282
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1282

```

Query Match 9.2%; Score 182.8; DB 11; Length 1734;

Best Local Similarity 49.2%; Pred. No. 3.1e-43;

Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

```

QY 79 TCTGCCCAACGACTTTGAATGGGCTTGGCAAGCGCCCTACAGATCCGAAGCGCCGT 138
DB 90 TTTTCTCGATGATTTTCATCTTTGGACAGCTGCTCGGCTTTGATGATGAGATGCAAC 149
QY 139 CAAGAAGGTGGCGCGCGCCGCTCATCTGGGACACGTACTGCACCTGAGCCATCGCG 198
DB 150 AAGTGAAGGTGGCAAGTCTTCAACTATATGGATCTTCAAGCTTCAAGTATCCAGAAAG 209
QY 199 CACCAACGCGCCCAACGCGAGATGGCTTGGCATCTACCAACGCTTACATGAGAGATT 258
DB 210 GACCAAAATGATATGAGATGAGATGAGCAATGATTTTATCATGCTTCAAGAGATGACAT 269
QY 259 TGAATCTCTTACCAAGTACGCGCGCAAGCGCTTCCGCTTCTCTGCTGCTGCTGCTGCT 318
DB 270 AAAATTGATGAGAGGCTTAAACATGAGAGCTTTCCGATTTTCAATCTGCTGCTGCTGCT 329
QY 319 CATTCCCTCGCGCGGCAAGCTGATCCGTCACAGAGAGGAAATTGATTTTACAGCA 378
DB 330 AATACCGAGTGAAGATGATGAAGATGAGTAAAGAGATGATGATGATGATGATGATGATG 389
QY 379 ACTGATGACGCGCTGTTGAGCGGCGGTATCAAGCTTGGGATCTTGTACCTACCTGGA 438
DB 390 TCTCATGACGAACTTTGCTTGTATGATGATCAATCAACCTTCGATGACGCTCTATCATTTGGA 449
QY 439 TCTGCTCAGGCGCTTCAAGATCGCTATGAGGCGTGGCTCAACGTAAGAGATCCAGCT 498
DB 450 CCACCCCAATTTTGGAGAGAGAAATGAGTGGCTTTCTAAGCCTTAATATGTA--- 506
QY 499 GGAATTTGAGCGGTATGAGAGTGTGCTTGAACGTTTGGGGAACGATCCAGACTG 558
DB 507 AGATTTTGAAGATTTTGAAGATTTGTTTGAAGATTTGGAAGATTTGAAGATTTG 566
QY 559 GATCAACATCAACGAGCCCTGATTCAGGCAATCTATGATATGCAACCGGAGCAAGCC 618
DB 567 GACAACATCAACGAGCCCTTATATATGATGATGCTGCGGTATGATCAAGGTAAACAGGC 626
QY 619 CCGCGGAGAGGCG-----AGCATTAACAAGCACTCAACCGAGGCAACACTGCACTGA 672
DB 627 GGTGGAAGATGCTCAAAATGGGTAAACGAAAGTGTGAGGCTGAGATTCAGATCCGA 686
QY 673 GCGGTGCTCGCTGGAAGGCGCCAGATCATGAGCCATGCGCGCGTGGCGGTCTACAG 732
DB 687 GCTTACATGTTTTCATACATCACTCTTGTGCCATGCGCTGCAAGTGAAGATTTG 746
QY 733 CAGGAGCTTTGCGCGCTGCAAAAGGCGCAATCGGCATCTGCTCAACGAGGACTACTA 792

```

```

DB 747 AAAATGTGAAAAAATCTTCGATGATGCGCAAAATTTGGATATGATATCAACCAAGATGTT 806
QY 793 TGAGCCCTGGGAGACGAAATGAGCTTCGGGACAGAGAGCTGTGAGCGACGATGAAATT 852
DB 807 CGAGCCTTATCATTTCCGATTTCAACTGACATTAAGAGAGCTGAGACGATCTTGCTT 866
QY 853 TCACATTTGCTGTTTGGCAATCCATCTTTGAAAGAGACTATCCAGAGCATGAA 912
DB 867 TGAATTTGATGCAATCTTATTCAGTCAATTC---ACGGATTTATTCAGAGATTTGAA 923
QY 913 GAAGCACTGGGAGAGAGGTTCCAGCCCTCACTCCCGGAGACTTTGGCATCTCAATGC 972
DB 924 AAGTACGCGGAAATTAATTAATCTTATTTACTGTTGAACAAATGATTTACAAA 983
QY 973 CGGAGAGACCACTTCTACGCGCATGATTAATCAATCCAGTTGCGCGCCACCTAGA 1032
DB 984 TTCA---TCAGATTTGCTTGAATTTATTTACTACAGCGCTCGCTGCTCATCTTCC 1040
QY 1033 CGGTCCCGTCCCGAGACGAGCATATCTGGCGGCATTCATGA-----GCAACAGGA 1083
DB 1041 TCACATGACCTCGAATAAACCCTGTTCAAACTGACACACATGTGGAATGAAATGAC 1100
QY 1084 GAATAGAGAGCGAGCGCCGTTGGCGAGAGAGCGGCTGCGCTGCGCTGCTGCTGCTGCTG 1143
DB 1101 TAAATCAGATGTCTACATTCAGACCTGGGAAAGAGAGGCTTTCTTATTTTCAACCC 1160
QY 1144 GGAATGTTCCGGAAGCATCTGCGCGGTGTACGCGCTGTAC---GGCAAGCCCATCTA 1200
DB 1161 GGAAGCTTACGAAAGTTTCTTAACTATATTAAGAGATTAATTAATGACATGCAAGCTA 1220
QY 1201 CATCACGGAAGAGATGCGCGCTGCGCTGCGAGAGAGAAATGACGTCGAGAGAGGCGCT 1260
DB 1221 CATCAAGAAATGAAATTAATGACAAAGAGAGCTTCAAAACCAAGAGAAATTTGT 1280
QY 1261 CAAGACCCCTTCGCAATCTGCGGTGTACCTGCACTTGAATGCACTTGAATTTCAAGGCT 1320
DB 1281 GAAGGACACATTTGATGATTAATTAATTAAGACATTTTGAAGAACTTATCAAAAGCTAT 1340
QY 1321 TACCCAGAGAGCGCTGCTGCTCAAGGGTACTTTGCGTGGCGCTTGTCTGATTAATCTTTGA 1380
DB 1341 AGTGAAGATGCGGTGACGTAAGAGATTTTACGATGATGATGATGATGATGATGATGAT 1400
QY 1381 ATGTCGATGCTTACGAGACCCAGATTCGG 1410
DB 1401 ATGGAGCATGATTAACCTGCAAGATTTGG 1430

```

Search completed: March 29, 2004, 03:16:31  
Job time : 706.681 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 23:31:59 ; Search time 6860.46 Seconds  
(without alignments)  
10211.501 Million cell updates/sec

Title: US-10-026-140-1  
Perfect score: 1991  
Sequence: 1 agccagtcgcacacgagcag.....ttttaaaaaaaaaaaaaa 1991

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_NA\_Main:\*

```
1: /cgn2_6/ptodata/2/pna/US081_PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
```

```
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES



QY 1021 GCGCCACTAGACGGTCCCGTCCCGAGAGAGCACTATCTCGGCGCCATCCATGAGACCA 1080  
DB 1021 GCGCCACTAGACGGTCCCGTCCCGAGAGAGCACTATCTCGGCGCCATCCATGAGACCA 1080  
QY 1081 GAGAGATAGAGACGGCAGCCCGTGGCGAGAGAGCGCCCTCGCTGGCTGCGCTCTG 1140  
DB 1081 GAGAGATAGAGAGCGCGCCCGTGGCGAGAGAGCGCCCTCGCTGGCTGCGCTCTG 1140  
QY 1141 CCGGAGCATGTCGGAGAGCATCTCGCCGGGTGTAGGGCTGTAGCGGCAAGCCACTTA 1200  
DB 1141 CCGGAGCATGTCGGAGAGCATCTCGCCGGGTGTAGGGCTGTAGCGGCAAGCCACTTA 1200  
QY 1201 CATACCGAGAGAGAGATGCGCGTCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 CATACCGAGAGAGAGATGCGCGTCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 CAAGAGACCCCTTCGCGATCCGCGTACTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 CAAGAGACCCCTTCGCGATCCGCGTACTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 TACCCAG 1380  
DB 1321 TACCCAG 1380  
QY 1381 ATGGTGAAGATGCTAG 1440  
DB 1381 ATGGTGAAGATGCTAG 1440  
QY 1441 GCGCAGCGCCAG 1500  
DB 1441 GCGCAGCGCCAG 1500  
QY 1501 TAAAGTGGCGGCAAG 1560  
DB 1501 TAAAGTGGCGGCAAG 1560  
QY 1561 TCTCTCTTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1620  
DB 1561 TCTCTCTTTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1620  
QY 1621 CCGCCGCGCTTCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
DB 1621 CCGCCGCGCTTCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
QY 1681 GCGCTTATACGCTTATAGCGCTGCTGTTACTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
DB 1681 GCGCTTATACGCTTATAGCGCTGCTGTTACTGCTGCTGCTGCTGCTGCTGCTGCT 1740  
QY 1741 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
DB 1741 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
QY 1801 AGCAGCAAGTTCAAG 1860  
DB 1801 AGCAGCAAGTTCAAG 1860  
QY 1861 AAGAGCCCATGTTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920  
DB 1861 AAGAGCCCATGTTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920  
QY 1921 CCGAGGAG 1980  
DB 1921 CCGAGGAG 1980  
QY 1981 AAAAAAAAAA 1991  
DB 1981 AAAAAAAAAA 1991

RESULT 2  
PCT-US02-34764-1  
; Sequence 1, Application PC/TUS0234764

GENERAL INFORMATION:  
APPLICANT: Dunn-Coleman, Nigel  
APPLICANT: Goedegebuur, Frits  
APPLICANT: Ward, Michael  
APPLICANT: Yao, Jian  
TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids  
FILE REFERENCE: GC697  
CURRENT APPLICATION NUMBER: PCT/US02/34764  
CURRENT FILING DATE: 2003-10-30  
NUMBER OF SEQ. ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ. ID NO 1  
LENGTH: 1991  
TYPE: DNA  
ORGANISM: Trichoderma reesei  
PCT-US02-34764-1

Query Match 100.0%; Score 1990.2; DB 1; Length 1991;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1991; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCAGGTGCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
DB 1 AGCCAGGTGCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
QY 61 AATGCCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 61 AATGCCGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
QY 121 CCAGATGAGAGGCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
DB 121 CCAGATGAGAGGCGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
QY 181 CCACTGAGAGCATCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 181 CCACTGAGAGCATCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
QY 241 CCGCTAGATGAG 300  
DB 241 CCGCTAGATGAG 300  
QY 301 CTTGCTGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
DB 301 CTTGCTGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
QY 361 AATTGAGTTTACAGCAAACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 361 AATTGAGTTTACAGCAAACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 421 GACTTTGATCAATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
DB 421 GACTTTGATCAATGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
QY 481 CTTGAGAGAGAGTCACTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 481 CTTGAGAGAGAGTCACTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
QY 541 GAGCCGAGTCCAGAGATGATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 541 GAGCCGAGTCCAGAGATGATCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 601 TGGCAGCGAG 660  
DB 601 TGGCAGCGAG 660  
QY 661 CACTGCACTGAG 720  
DB 661 CACTGCACTGAG 720  
QY 721 GCGCGTCTACAG 780  
DB 721 GCGCGTCTACAG 780

QY 781 CGGCGACTACTATGAGCCCTGGGACAGCATGATGAGCTGGGACAAAGAGCGCTGCTGACG 840  
DB 781 CGGCGACTACTATGAGCCCTGGGACAGCATGATGAGCTGGGACAAAGAGCGCTGCTGACG 840  
QY 841 ACGGATGGAATTTTCATTTGGCTGGTTTGGCAATCCATCTTCTTGAAGAAGACTATCC 900  
DB 841 ACGGATGGAATTTTCATTTGGCTGGTTTGGCAATCCATCTTCTTGAAGAAGACTATCC 900  
QY 901 AGAGGAGCATGAAAGAGCAGCTGGGGGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGC 960  
DB 901 AGAGGAGCATGAAAGAGCAGCTGGGGGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGC 960  
QY 961 CATCTCTCATGCGCGGAGAGAGCGAATCTTACGGGAGTAAATTAACAATCATCCAGTTCCG 1020  
DB 961 CATCTCTCATGCGCGGAGAGAGCGAATCTTACGGGAGTAAATTAACAATCATCCAGTTCCG 1020  
QY 1021 GCGCCACCTTAGACGGTCCCGTCCCGGAGACGGAATATCTCGGCGCATCATGAGACCA 1080  
DB 1021 GCGCCACCTTAGACGGTCCCGTCCCGGAGACGGAATATCTCGGCGCATCATGAGACCA 1080  
QY 1081 GGAAGATTAAGACGCGGACGCGCTGGGGAGAGAGCGGCTCCGCTGGCTGCGCTCTG 1140  
DB 1081 GGAAGATTAAGACGCGGACGCGCTGGGGAGAGAGCGGCTCCGCTGGCTGCGCTCTG 1140  
QY 1141 CCGGAGCATGTTCCGGAAGCATCTCGCCGGGTGAGCGGCTGTAACGGCAAGCCCATCTA 1200  
DB 1141 CCGGAGCATGTTCCGGAAGCATCTCGCCGGGTGAGCGGCTGTAACGGCAAGCCCATCTA 1200  
QY 1201 CATCAACGGAAGACGATGCCGCTGCTGAGAGAGAAACATGACGTGCGAGAGGCGCT 1260  
DB 1201 CATCAACGGAAGACGATGCCGCTGCTGAGAGAGAAACATGACGTGCGAGAGGCGCT 1260  
QY 1261 CAAGCAAGCCCTTCCGCAATCCGCTTGAATCTGCACTTGAATCTGCAATTTCAAGGCTAT 1320  
DB 1261 CAAGCAAGCCCTTCCGCAATCCGCTTGAATCTGCACTTGAATCTGCAATTTCAAGGCTAT 1320  
QY 1321 TACCCAGGACGCGCTGCTGCAAGGGGTACTTGGCGGGGCTGTCGATTAACCTTGA 1380  
DB 1321 TACCCAGGACGCGCTGCTGCAAGGGGTACTTGGCGGGGCTGTCGATTAACCTTGA 1380  
QY 1381 ATGCTCAGATGCTACGGAACCCAGATTCGGGCTGACGTTCAAGACTAACACCTCTCA 1440  
DB 1381 ATGCTCAGATGCTACGGAACCCAGATTCGGGCTGACGTTCAAGACTAACACCTCTCA 1440  
QY 1441 GCGCAGCGCCAAAGAGTCTGCTGCTCAAGGACATGTTGGCGCGCGGAGAGGT 1500  
DB 1441 GCGCAGCGCCAAAGAGTCTGCTGCTCAAGGACATGTTGGCGCGCGGAGAGGT 1500  
QY 1501 TAAAGTGGCGGATTAAGAAAGGAAATTTCTTCTGCAATTCAGCTCTATGACTTTCC 1560  
DB 1501 TAAAGTGGCGGATTAAGAAAGGAAATTTCTTCTGCAATTCAGCTCTATGACTTTCC 1560  
QY 1561 TCTCTCTTTTCCCTCCCTCCCTGTCCTCTCTCTCACTCTCAATTCCTCTATA 1620  
DB 1561 TCTCTCTTTTCCCTCCCTCCCTGTCCTCTCTCTCTCACTCTCAATTCCTCTATA 1620  
QY 1621 CCCCCCGCTTCTTCAATGACCCCATGCTGCTGCGCTGCGCTCTGCTGCAATTTCT 1680  
DB 1621 CCCCCCGCTTCTTCAATGACCCCATGCTGCTGCGCTGCGCTCTGCTGCAATTTCT 1680  
QY 1681 GCTCTTATCAACGCTTATGAGCTGCTTATCTTGGCTTTTCTTTTCTTTCTCTCTG 1740  
DB 1681 GCTCTTATCAACGCTTATGAGCTGCTTATCTTGGCTTTTCTTTTCTTTCTCTCTG 1740  
QY 1741 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
DB 1741 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
QY 1801 AGCACAAGTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1860  
DB 1801 AGCACAAGTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1860

QY 1861 AAAAGCCCATGTTTAAAGCTTCAATGCTATGTTTATCTTACATCAGCCATTCATA 1920  
DB 1861 AAAAGCCCATGTTTAAAGCTTCAATGCTATGTTTATCTTACATCAGCCATTCATA 1920  
QY 1921 CCAAGCAGAGGCGCAGAGAGTCTCATGCTTACCTGATATAGCTTTTAAAAA 1980  
DB 1921 CCAAGCAGAGGCGCAGAGAGTCTCATGCTTACCTGATATAGCTTTTAAAAA 1980  
QY 1981 AAAAAAAAAA 1991  
DB 1981 AAAAAAAAAA 1991

RESULT 3  
US-10-026-140-1  
; Sequence 1, Application US/10026140  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OR INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids  
; FILE REFERENCE: GC697  
; CURRENT APPLICATION NUMBER: US/10/026,140  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1991  
; TYPE: DNA  
; ORGANISM: Trichoderma reesei  
US-10-026-140-1

Query Match 100.0%; Score 1990.2; DB 43; Length 1991;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1991; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCAGGTGGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
DB 1 AGCCAGGTGGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 60  
QY 61 AATGCCGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
DB 61 AATGCCGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
QY 121 CCAAGTGAAGGCGCGTCAAGAAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 180  
DB 121 CCAAGTGAAGGCGCGTCAAGAAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 180  
QY 181 CCAAGTGAAGGCGCGTCAAGAAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 240  
DB 181 CCAAGTGAAGGCGCGTCAAGAAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTA 240  
QY 241 CCGCTACGATGAGGATTTGATCTCTTGAACCAAGTACGCGGCAAGAGGCTTACCGCTTC 300  
DB 241 CCGCTACGATGAGGATTTGATCTCTTGAACCAAGTACGCGGCAAGAGGCTTACCGCTTC 300  
QY 301 CTGTCGTGCTGCGGATCATTTCCCTCGCGCGGAGGCTGATCCGCTCAACGAGAGGG 360  
DB 301 CTGTCGTGCTGCGGATCATTTCCCTCGCGCGGAGGCTGATCCGCTCAACGAGAGGG 360  
QY 361 AATGAGTTTACAGAACTGATGAGCGCGCTGTTGAGCGGGGATATCAAGCTTGGGT 420  
DB 361 AATGAGTTTACAGAACTGATGAGCGCGCTGTTGAGCGGGGATATCAAGCTTGGGT 420  
QY 421 GACTTGTACCACTGGGATCTGCTCAGCGGCTTCAAGATCCGCTATGAGAGCTGCTCAA 480  
DB 421 GACTTGTACCACTGGGATCTGCTCAGCGGCTTCAAGATCCGCTATGAGAGCTGCTCAA 480  
QY 481 CGTGAAGAGGTCACGCTGGAATTTGAGCGGGATGCGAGGTTGCTTGAACGTTTGG 540  
DB 481 CGTGAAGAGGTCACGCTGGAATTTGAGCGGGATGCGAGGTTGCTTGAACGTTTGG 540

```

Qy 541 GAGCCGAGTCCAGAACTGATACATCAAGAAWCCCTGATTCAGGCCATCTATGATA 600
Db 541 GAGCCGAGTCCAGAACTGATACATCAAGAAWCCCTGATTCAGGCCATCTATGATA 600
Qy 601 TGCCACCGGAGCAACGCCCCGGGAGAGAGCAGATTAAAGAGCTTCAACGAGGCA 660
Db 601 TGCCACCGGAGCAACGCCCCGGGAGAGAGCAGATTAAAGAGCTTCAACGAGGCA 660
Qy 661 CACTGCACTGAGAGCCGTGCTGCTGAGAAAGGCCCAATCATAGGCCATGCCCCGCT 720
Db 661 CACTGCACTGAGAGCCGTGCTGCTGAGAAAGGCCCAATCATAGGCCATGCCCCGCT 720
Qy 721 GGCCTGTACAGAGAGGAGCTTTGCCCCCTGCAAAAGGCGCAGATCGGCACTTCTCAA 780
Db 721 GGCCTGTACAGAGAGGAGCTTTGCCCCCTGCAAAAGGCGCAGATCGGCACTTCTCAA 780
Qy 781 CGGCGACTACTATAGAGCCTTGAGACAGCAATGAGCCTCGGACAAAGAGGCTGCTGAG 840
Db 781 CGGCGACTACTATAGAGCCTTGAGACAGCAATGAGCCTCGGACAAAGAGGCTGCTGAG 840
Qy 841 ACCGATGGAATTTACATTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAAGACTATCC 900
Db 841 ACCGATGGAATTTACATTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAAGACTATCC 900
Qy 901 AGAGAGCATGAAAGAGAGCTGGGAGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGC 960
Db 901 AGAGAGCATGAAAGAGAGCTGGGAGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGC 960
Qy 961 CATCTCTAATGCGCGAGAGACCACTTCTACGGCATTAATTTACTACATCCAGTTCCG 1020
Db 961 CATCTCTAATGCGCGAGAGACCACTTCTACGGCATTAATTTACTACATCCAGTTCCG 1020
Qy 1021 GCGCCACTTACAGAGCTGCTCCGCTCCCGAGAGAGCAATCTGCGGCCCATCCATAGACCA 1080
Db 1021 GCGCCACTTACAGAGCTGCTCCGCTCCCGAGAGAGCAATCTGCGGCCCATCCATAGACCA 1080
Qy 1081 GAGAAATTAAGAGAGGAGCCCGCTTGAGAGAGAGGCGCTCGCTGCGCTGCGCTCTG 1140
Db 1081 GAGAAATTAAGAGAGGAGCCCGCTTGAGAGAGAGGCGCTCGCTGCGCTGCGCTCTG 1140
Qy 1141 CCGGAGCATGTTCCGAAAGCATCTCGCCGGGTGTACGGCTGTACGCGCAAGCCCATCTA 1200
Db 1141 CCGGAGCATGTTCCGAAAGCATCTCGCCGGGTGTACGGCTGTACGCGCAAGCCCATCTA 1200
Qy 1201 CATCAACGAGAGAGGATGCGCTGCTGAGAGAGAGAACTGACGTGCGAGAGGCGCT 1260
Db 1201 CATCAACGAGAGAGGATGCGCTGCTGAGAGAGAGAACTGACGTGCGAGAGGCGCT 1260
Qy 1261 CAAGGAGCCCTTCCGATCCGATCTTGACTGCGCATTTGGAATTCGAAGGCGCAT 1320
Db 1261 CAAGGAGCCCTTCCGATCCGATCTTGACTGCGCATTTGGAATTCGAAGGCGCAT 1320
Qy 1321 TACCAGAGAGCGGCTGCTGCTCAAGGGTACTTTGCGTGGCGCTTGTCTGATTAATTTGA 1380
Db 1321 TACCAGAGAGCGGCTGCTGCTCAAGGGTACTTTGCGTGGCGCTTGTCTGATTAATTTGA 1380
Qy 1381 ATGGTCAAGATGCTGCTGAGAGCCAGATTCGGGCTGACGTTCAAGACTTACACCTTCAA 1440
Db 1381 ATGGTCAAGATGCTGCTGAGAGCCAGATTCGGGCTGACGTTCAAGACTTACACCTTCAA 1440
Qy 1441 GCGCAGCGCCCAAGAAAGTGTGCTGCTGCTCAAGGAGCATGTTTGGCGCGCGAGAGG 1500
Db 1441 GCGCAGCGCCCAAGAAAGTGTGCTGCTGCTCAAGGAGCATGTTTGGCGCGCGAGAGG 1500
Qy 1501 TAAAGTGGCGCATTAAGAAAGGAAATTTCTTCTTGCATTTAGCATCTTATGCACTTCC 1560
Db 1501 TAAAGTGGCGCATTAAGAAAGGAAATTTCTTCTTGCATTTAGCATCTTATGCACTTCC 1560
Qy 1561 TCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATA 1620
Db 1561 TCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATA 1620

```

```

Qy 1621 CCCCCGCTTCTTCTCATAGAACCCCATGCTCTTCCCTTG6CCCTCTCTGTGCAATCT 1680
Db 1621 CCCCCGCTTCTTCTCATAGAACCCCATGCTCTTCCCTTG6CCCTCTCTGTGCAATCT 1680
Qy 1681 GCTCTTATACAGCTTTATGCTGCTGTTTACTTGGCTTTTCTTTTCTCTTCTGT 1740
Db 1681 GCTCTTATACAGCTTTATGCTGCTGTTTACTTGGCTTTTCTTTTCTCTTCTGT 1740
Qy 1741 CTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 1741 CTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Qy 1801 AGCACAAGTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1860
Db 1801 AGCACAAGTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1860
Qy 1861 AAAGCCCATGTTTAAAGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTT 1920
Db 1861 AAAGCCCATGTTTAAAGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTT 1920
Qy 1921 CAGGCGAAGAGGCGCAGAGAGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATG 1980
Db 1921 CAGGCGAAGAGGCGCAGAGAGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATG 1980
Qy 1981 AAAAAAAAAA 1991
Db 1981 AAAAAAAAAA 1991

```

```

RESULT 4
PCT-US02-34674-3
; Sequence 3, Application PC/RTU0234674
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuer, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: PCT/US02/34674
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
PCT-US02-34674-3

```

```

Query Match 73.0%; Score 1454.2; DB 1; Length 1455;
Best Local Similarity 100.0%; Pred. No. 2,2e-300;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 62 ATGCCGAGTCGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 121
Db 1 ATGCCGAGTCGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Qy 122 CAGATCGAAGGCGCCGTCGAAGAGGTGCGCGGCCGCTGCTCATCTTGGAGACGTAATG 181
Db 61 CAGATCGAAGGCGCCGTCGAAGAGGTGCGCGGCCGCTGCTCATCTTGGAGACGTAATG 120
Qy 182 CACTGAGAGCATGCGGCAACCAAGCGCGCAAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCT 241
Db 121 CACTGAGAGCATGCGGCAACCAAGCGCGCAAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCT 180
Qy 242 CGCTACGATGAGACTTTGATCTCTTGAACCAATGAGCGCAAGGCTTACGCTTCTCC 301
Db 181 CGCTACGATGAGACTTTGATCTCTTGAACCAATGAGCGCAAGGCTTACGCTTCTCC 240
Qy 302 TTGTCTGTGTCGCGGATCATTTCCCTTCGCGGAGGCTGATCCCTGCAACGAGAGGA 361
Db 241 TTGTCTGTGTCGCGGATCATTTCCCTTCGCGGAGGCTGATCCCTGCAACGAGAGGA 300

```

QY 362 ATTGAGTTTACAGCAAACTGATTTGAGCCCTGTTGAGGGGGGTATCAGCCCTTGGGTG 421  
|  
|  
Db 301 ATTGAGTTTACAGCAAACTGATTTGAGCCCTGTTGAGGGGGGTATCAGCCCTTGGGTG 360  
|  
QY 422 ACTTTGTAACCACTGGGATCTGCTCAGGCGCTTACAGATCGTATGGAAGGCTGGCTCAAC 481  
|  
Db 361 ACTTTGTAACCACTGGGATCTGCTCAGGCGCTTACAGATCGTATGGAAGGCTGGCTCAAC 420  
|  
QY 482 GTGGAAGAGGTTCAGCGCTGGACTTTGAGCGGTATGCCAGGTGTGCTTTGAAACGTTTTGGG 541  
|  
|  
Db 421 GTGGAAGAGGTTCAGCGCTGGACTTTGAGCGGTATGCCAGGTGTGCTTTGAAACGTTTTGGG 480  
|  
QY 542 GACCGAGTCCAGAACTGATCACCATCAAGAACCCCTGATTCAGGCACTTATGATAT 601  
|  
|  
Db 481 GACCGAGTCCAGAACTGATCACCATCAAGAACCCCTGATTCAGGCACTTATGATAT 540  
|  
QY 602 GCCACCGGACGAAACGCGCGGACAGAGAGCATTTAACAAGCACTTCCACGAGGGCAAC 661  
|  
|  
Db 541 GCCACCGGACGAAACGCGCGGACAGAGAGCATTTAACAAGCACTTCCACGAGGGCAAC 600  
|  
QY 662 ACTGCCACTGAGCCGTGCTCGCTGAAAAGCCAGATATAGCCATGCGCCGCGCTG 721  
|  
|  
Db 601 ACTGCCACTGAGCCGTGCTCGCTGAAAAGCCAGATATAGCCATGCGCCGCGCTG 660  
|  
QY 722 GCCGTCTACAGCAGGGACCTTTGCGCCCTCGCAAAAAGGCGCAGATCGCATCTCGCTCAAC 781  
|  
|  
Db 661 GCCGTCTACAGCAGGGACCTTTGCGCCCTCGCAAAAAGGCGCAGATCGCATCTCGCTCAAC 720  
|  
QY 782 GGCGACTATGAGCCCTGAGGACAGCAATGAGCCCTCGGACCAAGAGGCTGCTGAGCGA 841  
|  
|  
Db 721 GGCGACTATGAGCCCTGAGGACAGCAATGAGCCCTCGGACCAAGAGGCTGCTGAGCGA 780  
|  
QY 842 CGGATGGAATTTACATTTGCTGTTTGGCAATCCCATCTTTGAAGAGGACTATCA 901  
|  
|  
Db 781 CGGATGGAATTTACATTTGCTGTTTGGCAATCCCATCTTTGAAGAGGACTATCA 840  
|  
QY 902 GAGAGATGAAGAACAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGGC 961  
|  
|  
Db 841 GAGAGATGAAGAACAGAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGGC 900  
|  
QY 962 ATCTCTCAATGCCGAGAGACCGACTTCTACGCGCATGAATTACTACACATCCCACTGGCG 1021  
|  
|  
Db 901 ATCTCTCAATGCCGAGAGACCGACTTCTACGCGCATGAATTACTACACATCCCACTGGCG 960  
|  
QY 1022 CGGCACTGAGAGAGGTCCTCCCGAGACGAGCTATCTCGGCGCATTCAGACACAG 1081  
|  
|  
Db 961 CGGCACTGAGAGAGGTCCTCCCGAGACGAGCTATCTCGGCGCATTCAGACACAG 1020  
|  
QY 1082 GAGAAATGAAGACGAGCGCCGTTGGAGAGAGCGCCCTCGCTGGCTGCTCTGCTC 1141  
|  
|  
Db 1021 GAGAAATGAAGACGAGCGCCGTTGGAGAGAGCGCCCTCGCTGGCTGCTCTGCTC 1080  
|  
QY 1142 CCGGACATGTTCCGGAAGCATCTGCGCGGTGTATCGGCGCTGTATCGGCAAGCCCATCTAC 1201  
|  
|  
Db 1081 CCGGACATGTTCCGGAAGCATCTGCGCGGTGTATCGGCGCTGTATCGGCAAGCCCATCTAC 1140  
|  
QY 1202 ATCAACCGAGACGAGATGCGGTGCGCTGAGAGAGAAACATGAGCTGCGAGAGCGCTC 1261  
|  
|  
Db 1141 ATCAACCGAGACGAGATGCGGTGCGCTGAGAGAGAAACATGAGCTGCGAGAGCGCTC 1200  
|  
QY 1262 AAGCAACCCCTTCCGATCCGATCTTTGACTCGCATCTTGATCTGATTTCCAGAGGCCATT 1321  
|  
|  
Db 1201 AAGCAACCCCTTCCGATCCGATCTTTGACTCGCATCTTGATCTGATTTCCAGAGGCCATT 1260  
|  
QY 1322 ACCGAGAGCGCGCTGCTGCTCAAGGGGTACTTTGCGTGGGCGTGTCTCGATTAACCTTGA 1381  
|  
|  
Db 1261 ACCGAGAGCGCGCTGCTGCTCAAGGGGTACTTTGCGTGGGCGTGTCTCGATTAACCTTGA 1320  
|  
QY 1382 TGGTCAATGAGCTACGAGACCCAGATTGCGGCTCAAGCTTCAACAATCAACCCCTCAAG 1441  
|  
|  
Db 1321 TGGTCAATGAGCTACGAGACCCAGATTGCGGCTCAAGCTTCAACAATCAACCCCTCAAG 1380

QY 1442 CGCAGCCCAAGAAAGTCTGCCCTGATCCTCAAGGACATGTTTGCGGCCCGGACAGGGTT 1501  
|  
|  
Db 1381 CGCAGCCCAAGAAAGTCTGCCCTGATCCTCAAGGACATGTTTGCGGCCCGGACAGGGTT 1440  
|  
QY 1502 AAAGTGGCGGATTA 1516  
|  
|  
Db 1441 AAAGTGGCGGATTA 1455  
|  
RESULT 5  
PCT-US02-34764-3  
; Sequence 3 Application PC/TUS0234764  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids  
; TITLE OF INVENTION: Encoding the Same  
; FILE REFERENCE: GC697  
; CURRENT APPLICATION NUMBER: PCT/US02/34764  
; CURRENT FILING DATE: 2003-10-30  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1455  
; TYPE: DNA  
; ORGANISM: Trichoderma reesei  
PCT-US02-34764-3  
Query Match 73.0%; Score 1454.2; DB 1; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 2.2e-300;  
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 62 ATGCCGAGTGCCTAGCTCTGCGCAACGACTTTGATGAGGGGCTTGCAACGCGCGCTTAC 121  
|  
|  
Db 1 ATGCCGAGTGCCTAGCTCTGCGCAACGACTTTGATGAGGGGCTTGCAACGCGCGCTTAC 60  
|  
QY 122 CAGATGAAGAGGCGCGTCAAAAGAGTGGCCGCGCCGCTCATCTGGGACACGTAATGC 181  
|  
|  
Db 61 CAGATGAAGAGGCGCGTCAAAAGAGTGGCCGCGCCGCTCATCTGGGACACGTAATGC 120  
|  
QY 182 CACTTGAAGCATTCGCGACCAACGCGCGCAACGCGAGTGGCTTGGATCACTACAC 241  
|  
|  
Db 121 CACTTGAAGCATTCGCGACCAACGCGCGCAACGCGAGTGGCTTGGATCACTACAC 180  
|  
QY 242 CGCTAGAGTGAAGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCC 301  
|  
|  
Db 181 CGCTAGAGTGAAGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCC 240  
|  
QY 302 TTGTGCTGCTGCGGATCATTTCCCTCGGCGGCGAGGCTGATCCCGTCAACGAGAGGGA 361  
|  
|  
Db 241 TTGTGCTGCTGCGGATCATTTCCCTCGGCGGCGAGGCTGATCCCGTCAACGAGAGGGA 300  
|  
QY 362 ATTGAGTTTACAGCAAACTGATTCAGCCCTGTTGAGGCGGGGTATCAGCCCTTGGGTG 421  
|  
|  
Db 301 ATTGAGTTTACAGCAAACTGATTCAGCCCTGTTGAGGCGGGGTATCAGCCCTTGGGTG 360  
|  
QY 422 ACTTTGTAACCACTGGGATCTGCTCAGGCGCTTACAGATCGTATGGAAGGCTGGCTCAAC 481  
|  
|  
Db 361 ACTTTGTAACCACTGGGATCTGCTCAGGCGCTTACAGATCGTATGGAAGGCTGGCTCAAC 420  
|  
QY 482 GTGGAAGAGGTTCAGCGCTGGACTTTGAGCGGTATGCCAGGTGTGCTTTGAAACGTTTTGGG 541  
|  
|  
Db 421 GTGGAAGAGGTTCAGCGCTGGACTTTGAGCGGTATGCCAGGTGTGCTTTGAAACGTTTTGGG 480  
|  
QY 542 GACCGAGTCCAGAACTGATCACCATCAAGAACCCCTGATTCAGGCACTTATGATAT 601  
|  
|  
Db 481 GACCGAGTCCAGAACTGATCACCATCAAGAACCCCTGATTCAGGCACTTATGATAT 540  
|  
QY 602 GCCACCGGACGAAACGCGCGGACAGAGAGCATTTAACAAGCACTTCCACGAGGGCAAC 661  
|  
|  
Db 541 GCCACCGGACGAAACGCGCGGACAGAGAGCATTTAACAAGCACTTCCACGAGGGCAAC 600

```

QY 662 ACTGCCACTGAGCGGTGCTGCTGGAAAGGCCAGATCATGAGCCATGCGCCGCGTG 721
DB 601 ACTGCCACTGAGCGGTGCTGCTGGAAAGGCCAGATCATGAGCCATGCGCCGCGTG 660
QY 722 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCGCAGATCGGCATCTTCAC 781
DB 661 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCGCAGATCGGCATCTTCAC 720
QY 782 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTAAGCA 841
DB 721 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTAAGCA 780
QY 842 CGGATGAAATTTACATTTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 901
DB 781 CGGATGAAATTTACATTTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 840
QY 902 GAGAGCATGAAAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGACCTTTCG 961
DB 841 GAGAGCATGAAAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGACCTTTCG 900
QY 962 ATCTCTAATGCGCGAAGAGCCGACTTCTAGGCAATGAAATTAACAATCCAGTTTCGG 1021
DB 901 ATCTCTAATGCGCGAAGAGCCGACTTCTAGGCAATGAAATTAACAATCCAGTTTCGG 960
QY 1022 CGGCACCTAGAGCGGTCCCGTCCCGGAGAGGAGCAATCTCGGGCGCATTCAGAGCA 1081
DB 961 CGGCACCTAGAGCGGTCCCGTCCCGGAGAGGAGCAATCTCGGGCGCATTCAGAGCA 1020
QY 1082 GAGAATTAAGAGCGGACCCGCTTGGCGAGAGAGCGGCTTCGCTGCTGCTCTG 1141
DB 1021 GAGAATTAAGAGCGGACCCGCTTGGCGAGAGAGCGGCTTCGCTGCTGCTCTG 1080
QY 1142 CGGGAATGTTCCGGAAGACTCTCGCCCGGGGTGAGCGCTTGAACGCAAGCCATCTAC 1201
DB 1081 CGGGAATGTTCCGGAAGACTCTCGCCCGGGGTGAGCGCTTGAACGCAAGCCATCTAC 1140
QY 1202 ATCAGCGAGAGGAGTGGCCCGTCCGCGAGAGAGAGCAATGAGTGGCGAGAGGCGT 1261
DB 1141 ATCAGCGAGAGGAGTGGCCCGTCCGCGAGAGAGAGCAATGAGTGGCGAGAGGCGT 1200
QY 1262 AACGACCCCTTCGCAATCCGTAATCTTGAATCGCACTTGGACTGATTTCCAGAGCAT 1321
DB 1201 AACGACCCCTTCGCAATCCGTAATCTTGAATCGCACTTGGACTGATTTCCAGAGCAT 1260
QY 1322 ACCCAGAGCGGCGTCTGCTCAAGGGGTAATTTGCGTGGGCTTCTGATTAATTGGAA 1381
DB 1261 ACCCAGAGCGGCGTCTGCTCAAGGGGTAATTTGCGTGGGCTTCTGATTAATTGGAA 1320
QY 1382 TGGTCAGATGGCTACGGAACCCAGATTCGCGTCAAGCTTCAAGACTACACCCCTCAG 1441
DB 1321 TGGTCAGATGGCTACGGAACCCAGATTCGCGTCAAGCTTCAAGACTACACCCCTCAG 1380
QY 1442 CGCAGCCCAAGAGTCTGCGCTGCTCTCAAGAGCAATGTTTGGCGCCGCGAGAGGTT 1501
DB 1381 CGCAGCCCAAGAGTCTGCGCTGCTCTCAAGAGCAATGTTTGGCGCCGCGAGAGGTT 1440
QY 1502 AAAAGTGGCGCATTA 1516
DB 1441 AAAAGTGGCGCATTA 1455

```

```

RESULT 6
US-10-026-140-3
; Sequence 3, Application US/10026140
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697

```

```

; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-3

Query Match 73.0%; Score 1454.2; DB 43; Length 1455;
Best Local Similarity: 100.0%; Pred. No. 2.2e-300;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 ATGCCGAGTCCGTACTCTGCCCAACGACTTGAATGGGGCTTGGCAACGCGGCTTAC 121
DB 1 ATGCCGAGTCCGTACTCTGCCCAACGACTTGAATGGGGCTTGGCAACGCGGCTTAC 60
QY 122 CAGATGAAGGCGCGCTCAAGAGAGTGGCCGCGGCGCTTCATCTGGGACAGTACTGC 181
DB 61 CAGATGAAGGCGCGCTCAAGAGAGTGGCCGCGGCGCTTCATCTGGGACAGTACTGC 120
QY 182 CACTGAGCATCGCGACCAACGCGCCCAACGCGGATGCTTGGATCACTACAC 241
DB 121 CACTGAGCATCGCGACCAACGCGCCCAACGCGGATGCTTGGATCACTACAC 180
QY 242 CGCTAGAGAGCACTTGAATCTTGAACCAAGTACGCGCAAAAGGCTTACGCTTCC 301
DB 181 CGCTAGAGAGCACTTGAATCTTGAACCAAGTACGCGCAAAAGGCTTACGCTTCC 240
QY 302 TTGTCTGATCGCGATCATTTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGA 361
DB 241 TTGTCTGATCGCGATCATTTCCCTCGGCGGAGGCTGATCCGTCACAGAGAGGA 300
QY 362 ATTAGATTTTACAGCAACTGATTAAGCGCTGTTAGGCGGGGATACAGCTTGGTG 421
DB 301 ATTAGATTTTACAGCAACTGATTAAGCGCTGTTAGGCGGGGATACAGCTTGGTG 360
QY 422 ACTTTTACACTGGGATCTGCTCGAGGCTTCAAGATCGCATAGAGGCTGGCTAAC 481
DB 361 ACTTTTACACTGGGATCTGCTCGAGGCTTCAAGATCGCATAGAGGCTGGCTAAC 420
QY 482 GTGGAAGAGTCCAGTGAATTTGAGCGGTAATGAGGTTGCTTGAACCTTTGGG 541
DB 421 GTGGAAGAGTCCAGTGAATTTGAGCGGTAATGAGGTTGCTTGAACCTTTGGG 480
QY 542 GACCGAGTCCAGAACTGATCAACATCAAGACCTGGAATTCAGGCTCATATGATAT 601
DB 481 GACCGAGTCCAGAACTGATCAACATCAAGACCTGGAATTCAGGCTCATATGATAT 540
QY 602 GCCACCGGAGCAAGCCCGGGCAGAGACAGATTAACAAGACTCCACCGAGGGCAAC 661
DB 541 GCCACCGGAGCAAGCCCGGGCAGAGACAGATTAACAAGACTCCACCGAGGGCAAC 600
QY 662 ACTGCCACTGAGCGGTGCTGCTGGAAGAGCCAGATCATGAGCCATGCGCGCGTG 721
DB 601 ACTGCCACTGAGCGGTGCTGCTGGAAGAGCCAGATCATGAGCCATGCGCGCGTG 660
QY 722 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCGCAGATCGGCATCTTCAC 781
DB 661 GCCGTCTACAGCAGGAGCTTTCGCCCCCTCGCAAAAGGCGCAGATCGGCATCTTCAC 720
QY 782 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTAAGCA 841
DB 721 GGGGACTACTATAGAGCCCTTGGGACAGCAATGAGCTTCGGGACAAAGAGGCTGTAAGCA 780
QY 842 CGGATGAAATTTACATTTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 901
DB 781 CGGATGAAATTTACATTTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCA 840
QY 902 GAGAGCATGAAAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGACCTTTCG 961
DB 841 GAGAGCATGAAAGAGCAGTGGGCGAGAGGCTTTCAGCCCTCACTCCCGGACCTTTCG 900

```



QY 962 ATCTCAATGCGGAGAGACCGACTTCTACGGCATGTAATTACTACACATCCAGTTGGCG 1021  
Db 901 ATCTCAATGCGGAGAGACCGACTTCTACGGCATGTAATTACTACACATCCAGTTGGCG 960  
QY 1022 CGCCACCTAGACGGTCCCGTCCCGAGACGAGCTATCTCGCGCCCATCTCAAGACACAG 1081  
Db 961 CGCCACCTAGACGGTCCCGTCCCGAGACGAGCTATCTCGCGCCCATCTCAAGACACAG 1020  
QY 1082 GAGAAATAGACGGACCGCCGTTGGCGAGAGACGGCGCTCGCTGGCTGGCTCTGCG 1141  
Db 1021 GAGAAATAGACGGACCGCCGTTGGCGAGAGACGGCGCTCGCTGGCTGGCTCTGCG 1080  
QY 1142 CGGAGACGTTTCGGAGACATCTGCGCCGGGTGTAAGCGCCGTATCGGCAAGCCCATCTAC 1201  
Db 1081 CGGAGACGTTTCGGAGACATCTGCGCCGGGTGTAAGCGCCGTATCGGCAAGCCCATCTAC 1140  
QY 1202 ATCAACGAGAACGGATCCCGTGGCTGAGAGAGAGAACATGACGTGCGAGAGCGCGTC 1261  
Db 1141 ATCAACGAGAACGGATCCCGTGGCTGAGAGAGAGAACATGACGTGCGAGAGCGCGTC 1200  
QY 1262 AAGCAACCTTTCGGCATTCGCTACTTGAATCTGCACTTGAATCTGATTTCCAAAGCCATT 1321  
Db 1201 AAGCAACCTTTCGGCATTCGCTACTTGAATCTGCACTTGAATCTGATTTCCAAAGCCATT 1260  
QY 1322 ACCAGAGACGGCGTCTGCTCAAGGGGTACTTTGCGGGCGTTGCTGATTAATTGGA 1381  
Db 1261 ACCAGAGACGGCGTCTGCTCAAGGGGTACTTTGCGGGCGTTGCTGATTAATTGGA 1320  
QY 1382 TGGTCAGATGCTACCGAACCGAATTCGGCGCTCACTTCAAGACTACACCAACCTCAAG 1441  
Db 1321 TGGTCAGATGCTACCGAACCGAATTCGGCGCTCACTTCAAGACTACACCAACCTCAAG 1380  
QY 1442 CGCAGCGCCCAAGAGTCTGCTGCTCTCAAGAGACATGTTTGGCGCCCGCAAGAGGTT 1501  
Db 1381 CGCAGCGCCCAAGAGTCTGCTGCTCTCAAGAGACATGTTTGGCGCCCGCAAGAGGTT 1440  
QY 1502 AAAGTGGCGGATTA 1516  
Db 1441 AAAGTGGCGGATTA 1455

RESULT 7  
US-09-533-559-7458  
; Sequence 7458, Application US/09533559  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster  
; APPLICANT: Sakari Kauppinen  
; APPLICANT: Ib Groth Clausen  
; APPLICANT: Peter Bjarke Olsen  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 5849.200-US  
; CURRENT APPLICATION NUMBER: US/09/533,559  
; EARLIER FILING DATE: 2000-03-22  
; EARLIER APPLICATION NUMBER: 09/273,623  
; NUMBER OF SEQ ID NOS: 7860  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7458  
; LENGTH: 735  
; TYPE: DNA  
; ORGANISM: Tricoderma reesei  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(735)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-533-559-7458

Query Match 30.0%; Score 598; DB 22; Length 735;  
Best Local Similarity 97.0%; Pred. No. 2.9e-117;

Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;  
QY 319 CATTCCTCTGAGGAGAGCTGATCCCGTCAACGAGAGAGAAATGATTTACAGCAA 378  
Db 1 CATTCCTCTGAGGAGAGCTGATCCCGTCAACGAGAGAGAAATGATTTACAGCAA 60  
QY 379 ACTGATTAAGACGCCCTGTTAGAGCGGGGTATCAAGCCTTGGGTGACTTTGTACCACTGGGA 438  
Db 61 ACTGATTAAGACGCCCTGTTAGAGCGGGGTATCAAGCCTTGGGTGACTTTGTACCACTGGGA 120  
QY 439 TCTGCTCAGAGCGCTTCAAGATGCTATGAGAGCTGCTCAAGTGGAAAGATCCAGCT 498  
Db 121 TCTGCTCAGAGCGCTTCAAGATGCTATGAGAGCTGCTCAAGTGGAAAGATCCAGCT 180  
QY 499 GGACTTTAGCGGATAGCGAGTGTGCTTTGAAAGCTTTTGGGACCGAGTCCAGAACTG 558  
Db 181 GGACTTTAGCGGATAGCGAGTGTGCTTTGAAAGCTTTTGGGACCGAGTCCAGAACTG 240  
QY 559 GATCACCATCAACGAGCCCTGATTCAGAGCCATCTATGATATGCCACCGGACAGAAAGC 618  
Db 241 GATCACCATCAACGAGCCCTGATTCAGAGCCATCTATGATATGCCACCGGACAGAAAGC 300  
QY 619 CCGGGGAGAGAGCACTTAACAAGCACTCCAGAGGGAGAACTGGCCACTGAGCCGTG 678  
Db 301 CCGGGGAGAGAGCACTTAACAAGCACTCCAGAGGGAGAACTGGCCACTGAGCCGTG 360  
QY 679 GCTCGTGAAGAGCCAGATCATGAGCCATGCGCCGCGCGTGGCGCTTACAGCAGGGA 738  
Db 361 GCTCGTGAAGAGCCAGATCATGAGCCATGCGCCGCGCGTGGCGCTTACAGCAGGGA 420  
QY 739 CTTTGGCCCTCGCAAAAGGCGCAGATGCGCATCTGCTCAAGCGGACTATATGAGCC 798  
Db 421 CTTTGGCCCTCGCAAAAGGCGCAGATGCGCATCTGCTCAAGCGGACTATATGAGCC 480  
QY 799 CTGGAAGAGATGAGCTCTGGGACAAAGAGGCTGCTGAGCGAGATGGAATTTGCAT 858  
Db 481 CTGGAAGAGATGAGCTCTGGGACAAAGAGGCTGCTGAGCGAGATGGAATTTGCAT 540  
QY 859 TGGCTGTTGGCCATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGATGAAGAGC 917  
Db 541 TGGCTGTTGGCCATCCCATCTTCTTGAAGAAGACTA-TCCAGAGAGATGAAGAGC 600  
QY 918 AGCT--GGGCGAGAGGCTTCAGGCTTCAC-TCCCGGAGCTTT--GCCATCTCAATGC 972  
Db 601 AACTTGGGCGAGAGGCTTTCACCTTACTCCCGGAGCTTTTGCATCTTNAATGC 660  
QY 973 CGAGAGAGAC 982  
Db 661 CGAGAGAGAC 670

RESULT 8  
US-10-653-047-7458  
; Sequence 7458, Application US/10653047  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster  
; APPLICANT: Sakari Kauppinen  
; APPLICANT: Ib Groth Clausen  
; APPLICANT: Peter Bjarke Olsen  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE REFERENCE: 5849.200-US  
; CURRENT APPLICATION NUMBER: US/10/653,047  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US/09/533,559  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 09/273,623  
; NUMBER OF SEQ ID NOS: 7860  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7458



```

; LENGTH: 735
; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(735)
; OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7458

```

Query Match 30.0%; Score 598; DB 53; Length 735;

Best Local Similarity 97.0%; Pred. No. 2,9e-117;  
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

```

QY 319 CATTCCTCGCGCGAGCTGATCCCGTCAACGAGAGGAAATTGATTTTACAGCA 378
Db 1 CATTCCTCGCGCGAGCTGATCCCGTCAACGAGAGGAAATTGATTTTACAGCA 60
QY 379 ACTGATTGACGCCCTGTTGAGCGGGGATACGCCCTTGGGTGACTTTTACCACTGGGA 438
Db 61 ACTGATTGACGCCCTGTTGAGCGGGGATACGCCCTTGGGTGACTTTTACCACTGGGA 120
QY 439 TCTGCTCAGAGCGCTTACGATCGCTATGAGAGTGGCTCAACGTGGAAGAGTCCAGCT 498
Db 121 TCTGCTCAGAGCGCTTACGATCGCTATGAGAGTGGCTCAACGTGGAAGAGTCCAGCT 180
QY 499 GCACTTTGAGCGGATGATGAGAGTGTGCTTTGAAAGTTTGGGAGCCGATCCAGAACTG 558
Db 181 GCACTTTGAGCGGATGATGAGAGTGTGCTTTGAAAGTTTGGGAGCCGATCCAGAACTG 240
QY 559 GATCACCATTCAACGAGCCCTGATTCAGGCACTCTATGATGATGCAACCGGACGCAACGC 618
Db 241 GATCACCATTCAACGAGCCCTGATTCAGGCACTCTATGATGATGCAACCGGACGCAACGC 300
QY 619 CCGGCGAGAGAGCACTTAAACAAGCACTCAACGAGGCAACACTCTCCACTGAGCGGTG 678
Db 301 CCGGCGAGAGAGCACTTAAACAAGCACTCAACGAGGCAACACTCTCCACTGAGCGGTG 360
QY 679 GCTCGTGAAGAGCCGAGATCATGAGCACTGCGCGCGGCTGGCGCTTCAACAGCAGGA 738
Db 361 GCTCGTGAAGAGCCGAGATCATGAGCACTGCGCGCGGCTGGCGCTTCAACAGCAGGA 420
QY 739 CTTTCGCCCTCCCAAAAGGCGCAGATCGGCACTCTGCTCAACGCGCACTACTATGAGCC 798
Db 421 CTTTCGCCCTCCCAAAAGGCGCAGATCGGCACTCTGCTCAACGCGCACTACTATGAGCC 480
QY 799 CTGGGACAGAAATGAGCTTGGGCAAGAGAGCTGCTGAGCGAGATGGAATTTTCAT 858
Db 481 CTGGGACAGAAATGAGCTTGGGCAAGAGAGCTGCTGAGCGAGATGGAATTTTCAT 540
QY 859 TGAGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCGAGAGATGGAAGAAGC 917
Db 541 TGAGCTGTTGGCAATCCCATCTTCTTGAAGAAGACTA-TCCGAGAGATGGAAGAAGC 600
QY 918 AGCT--GGGCGAGAGGCTTCCAGGCTTCC-TCCCGGAGCTTT--GCCATCTCAATGC 972
Db 601 AGCTGGGGGAGAGGCTTCCAGGCTTCC-TCCCGGAGCTTT--GCCATCTCAATGC 660
QY 973 CGGAGAGACC 982
Db 661 CGGAGAGAAC 670

```

RESULT 9  
US-09-417-507-9340

```

; Sequence 9340, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUNGICIDUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312

```

```

; SEQ ID NO 9340
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1562)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-417-507-9340

```

Query Match 17.3%; Score 345.4; DB 20; Length 1572;

Best Local Similarity 55.7%; Pred. No. 3.8e-63;  
Matches 721; Conservative 2; Mismatches 563; Indels 9; Gaps 3;

```

QY 88 CGACTTTGAATGGGGCTTGCAACGCGCGCTCAACGATGAGAGGCGCGTCAAGAAAG 147
Db 66 CGACTTTTTCATGATTTACCGCACCGGGGCGCCAGGTGAAAGAGACTTGGAAATTAAGA 125
QY 148 TGGCGGCGCGCGCTTCATCTGGGACACGTACTGCCACTGGAGCCATCGGCACCAACGG 207
Db 126 CGGTAAAGGCCAATCAATCTGGGATCTTTTGGGCAACACCTGGGAAGGTGAAGAGCGG 185
QY 208 CGCCAAAGCGGATGTGGCTTGCATCACTACACCGCTACAGTAGAGACTTTGATCTTT 267
Db 186 AAGCAGCTGGGAGACGACGCGGTACGCTGTATGACCTGTACAGAAAGATGTGGCTCAT 245
QY 268 GACCAAGTACGGGCAAAAGGCTTACCGCTTCTGCTGTGGTGGCGATTCATTCCTCT 327
Db 246 GAAGTGTATGGGCTCAATGATCAACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 305
QY 328 CGGCGGAGGCTGATCCCGTCAACGAGAGGAAATTGATTTTACAGCAACTGATTTGA 387
Db 306 TGGAGGGGTGGAGAGACTCTGTCACAGAAAGGGGATTTGATTAATTAATTAATTTCTTT 365
QY 388 CGCCCTGTGAGCGGGGATACAGCCTTGGGATCTTTGATCACTCTGGATCTGCTCA 447
Db 366 CGAGCTTCTCCGTAATGAGCATCAACCGTTTGTACCTCTTCCACTGGGACACCGCA 425
QY 448 GGGGCTTCAACGATGATGAGAGCTGGCTCAACGTGGAAGAGGTCCAGTGAATTGA 507
Db 426 ATGCTGAGAGATGCTCAACGCGGCAATGCTGAACAGAAAGTTGTGCGGACTTTGT 485
QY 508 GCGGTATGCGAGGTTGTGCTTTTGAAGTTTGGGAGCCGATCCAGAACTGATACCAT 567
Db 486 CAATTAGCTCGTGTGTTGCTTGAAGAGACTCGGCGATCGGCTCAAGATTTGATCACT 545
QY 568 CAACGAGCCCTGATTCAGGCACTTATGAGATATGCCACCGGACGACCGCGGCGAG 627
Db 546 CAACGAGCCCTGATTCAGGCACTTATGAGATATGCCACCGGCGGCTCCACCGGCGCG 605
QY 628 GAGCAGATTAAACAAGCATCCACCGAGGCAACATGCCACTGAGGCGGCTCGCTGG 687
Db 606 GTCTTATTCGCGGACCGCAACGAGAGGCGGCTGCTCCACGAGCATTCATCTGTGC 665
QY 688 AAGAGCCAGATCATGAGCCATCCGCGCGCTGAGCGCTTACAGAGGACTTTGCGCC 747
Db 666 GCAAGAGAGCTGTGGCCACGAGGACGATGTCCTCTTCAAGAGAGTTCCAGCC 725
QY 748 CTGCAAAAGGCGCAATCGCATCTTCTCAACGCGGCACTACTATGAGCTTGGGACAG 807
Db 726 CCATCAGCAAGGAGCATCGGCATTCACGCTTACGCGGAATGTGTCGCAATGGGATGA 785
QY 808 CAATGAGCTTCGGGACCAAGAGGCTCTGAGCGACGAGATGGAATTTTCAATTTGGCTGT 867
Db 786 AGCCGAGCCGCTGATCAAGCGGCGCTGAGCGCGCGCGGAGATTTGAGATTCCTGTGT 845
QY 868 TGCCAAATCCCATCTTTTGAAGAAGACTATCAGAGAGCATGAAGAAGCACTGGGCGA 927
Db 846 CGGAGAGCGGCTTCAAGAGCGGATGATATCGGCGCTGATGAGAGCCAGCTGGGTGA 905
QY 928 GAGGCTTCCAGCCCTCACTCCGCGGACTTTGCAATCTCTCAATGCGGAGAGACCGAATT 987
Db 906 CGGCTGCTCCAAAGTCAACCGCCGAGAGAT---CAAAGCTGTGCTTGGAGAGCTTCGAGTT 962

```



```

Db      1447 GACCTGCTAGACGAGTACTCAATATATGAGATTACGCCATTCGTAACTCTTCACTGG 1506
Qy      437 GATGTGCTAGAGGGCTTCAAGATCGCTATAGAGGCTGCTCAACGAGGAAGAGTCCAG 496
Db      1507 GATGTTCGAGAGGTTTGGAGATCGGTACGGGGCATGTGTAACGAGAGAGTTTATT 1566
Qy      497 CTGAGCTTTAGAGCGTATGCGAGGTTTGCTTTGAAGCTTTTGGGAGCCGAGTCCAGAAC 556
Db      1567 CTGAGCTTTTCCGCTACGCGCGTGTCTGTTTGAAGCCTTAGGCGCAAGAGTCCGAC 1626
Qy      557 TGGATCAATCAACGAGCCCTGATTCAGGCGCATCTATGATATGTCACCGGAGCAAC 616
Db      1627 TGGATTAATCAATGAACCTGGGTGTACTCGCTGACAGGATATGCGGAGGTTTCAT 1686
Qy      617 GCGCCGGGCGAGAGCAGCATTTAACAGACATCCACCGAGGCAACATGCGCATGAGCCG 676
Db      1687 GCTCTGCGACGGTTCATGTTTCCGGAATCAATGAAGAGGTGATTCCTGACGAAACCG 1746
Qy      677 TGGCTGCTGGAAGGCCAGATCATGAGCCATGCGCGCGCTGCGCTTACAGCAGG 736
Db      1747 TTTATCGTGGCCACACAAACCTGTGTACCCAGGCGCATGTTTCCAAAGCTTTACCGAG 1806
Qy      737 GACTTTGCGCCCTCGCAAAAGGGCGAGATCGGCACTCTGCTCAACGCGGCACTACTATAG 796
Db      1807 GTGTTCCAAACCGAGAGAAAGAAACAATTGGCATCACTCCATGCGCACTGGTCCGAA 1866
Qy      797 CCTGCGGACAGCATGAGCTCTGGGACAAAGAGGCTGTGAGCAGCAGGATGGAATTTAC 856
Db      1867 CTTTGGAGTAAAGCATGCTCGCGGAGCAGAGAGCAGAAACGAGGCAAGGAAATTTGAG 1926
Qy      857 ATTGGCTGTTTCCAAATCCCATCTTCTTGAAGAGAGCTATCCAGAGAGCATGAAGAAG 916
Db      1927 ATGGCTTGTGTTCCGGAGCCCTTATTAACAAGCAGGCGCATACCCAGCTCTCAATGCGG 1986
Qy      917 CAGCTGGGCGAGAGGCTTCCAGCCCTGACTCCCGCGGACTTTGCGCATCTCAATGCGG 976
Db      1987 CAATCGGGAGCGGTTCACCGCGTTTCACTCCGAGAGAGT---CAAAATGTTGACTAGGA 2043
Qy      977 GAAACCGACTTTCAGGCAATGAATTAACAATCCAGTTTGGCGGCGCATAGACGCT 1036
Db      2044 AGTTCAAGAAATTTTACGGTATGAATCGTACACGACTTTTTCGTCAGCAGCAAGATACG 2103
Qy      1037 CCGGCTCCCGAGAGCGACTATCTCGGCGCCATCCATGAGCAGCAGAAATTAAGAGCGG 1096
Db      2104 CTTCCAGATATCAATGACCAAAAGGCAATGCTTTTTCACATTAACAAGAGAGAGG 2163
Qy      1097 AGCCCGGTTGGGAGAGAGCGGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1156
Db      2164 GTCTTCGAGGCGAAGATCGGATAGCGCTTGGCTGCGGAGCGGACCCATGAGATGGCG 2223
Qy      1157 AAGCATCTGCGCCCGGGGTGACGGCTGTACGGCAAGCCCATCTACATCAGCAGAGAGGA 1216
Db      2224 AAGCTCTCAATTTGATTTGAAACCGATATCATGTGCTTATATATGTCAGGAAATG-- 2281
Qy      1217 TGGCCGTCGCTGAGAGAGAGAAATGACGTCGAGAGAGGCGCGTCAACGACCCCTTCCG 1276
Db      2282 -GCACAGACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy      1277 ATCCGATCTTTGATCTGCACTTGAATGCAATTTCCAAAGGCAATTAACCAAGAGAGGCT 1336
Db      2341 ATGCGATTTCTTGAAGGCTAGTGTGTGATGTTGGCCGCTGTCTGTCAAAAGAGATGAGTG 2400
Qy      1337 GTGCTCAAGGGGATCTTTGCTGCTGGGCTTGTGCTGATTAACCTGG 1379
Db      2401 GATATCCGATCTTACTTGTGCTGAGACCTTCAACGAGCAACTGG 2443

```

RESULT 12  
 US-10-369-493-37092  
 ; Sequence 37092; Application US/10369493  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei

```

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIORITY FILING DATE: 2003-02-28
; PRIORITY FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37092
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-37092

Query Match      16.0%; Score 317.8; DB 50; Length 1470;
Best Local Similarity 53.8%; Pred. No. 3e-57;
Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

Qy      81 TGGCCAAAGACTTTGAATGAGGCTTGCGCAACGCGCCCTTACAGATGGAAGCGCGCTCA 140
Db      2 TGGCGAATGACTTCTTCATGATATGCGACGCGCGCGCAGGTTGAAGAGCGTGA 61
Qy      141 AAGAGGTGGCGCGCGCGCTGCATGCGGACAGTACCTGCACTGAGCATTGCGCA 200
Db      62 ACAAGATGGCAAGGCTCTTCATGAGTGGGACACATTCGGGCACTCCAGGCAAGTCA 121
Qy      201 CCAAGCGCGCAACGCGGATGTGCTTGCATCACTCAACCGCTACAGATGAGAGCTTTG 260
Db      122 AAGACATATCAATGCGGATGAGCTGTGAGTTTACAGCTTTTACCGGAAAGATTTG 181
Qy      261 ATCTCTTGAACAAATGAGCGCGCAAGGCTTACCGCTTCTCTTGTGTGCTGCGATCA 320
Db      182 CCTGATGAAGTGTATGTGTCAACGCTTATGATTCGTGCTATCTGCTGCGCATAA 241
Qy      321 TTTCCCTCGCGGCGAGGCTGATCCGCTCAACGAGAGGGAATGAGTTTACAGCAAC 380
Db      242 TTCACACTCGGAGCGCGGACGACCGGCTCAATAGCAAGGAACTCACTACAGAGAC 301
Qy      381 TGAATGAGCCCTGTTGAGGCGGGGTATCAAGCTTGGTGTGACTTTGTACCATGAGATC 440
Db      302 TGTGAGACAGATCACTCAATATGAGATTAAGCATTCGTATCCATCTTTCACCTGGATG 361
Qy      441 TGCCTAGGCGCTTACAGATGCTATGAGAGCTGAGCTCAAGTGAAGAGTCCAGCTG 500
Db      362 TTCGCGAGGCTTTGAGAGATCGGTACGGGGCATGTGAACAGAGAGGTTCAATTCCTG 421
Qy      501 ACTTTGAGCGGTATGCGAGTGTGCTTTTGAACGTTTGGGGACCGAGTCCAGAACTGGA 560
Db      422 ACTTTGCGCTACGCGCGGTGTGTTTGAAGCGCTTAAGGCGCAAGAGTCCGTACCTGA 481
Qy      561 TCACATCAACGAGMCCCTGTGATTCAAGGCACTATGATATGCAACCGGAGCAACGCC 620
Db      482 TTACATTTCAATGAACCTGTGTGTACTCGCTGCGAGGTTATGCGGAGAGTTCATCTC 541
Qy      621 CGGCGAGAGCAGCATTTAACAGCACTCAACGAGGCAACACTGCTCACTGAGCGCTGG 680
Db      542 CTGACAGGTCATGTTTCGGAATCAATGAAAGGAGTGAATTCCTGACAGAAACGTTCA 601
Qy      681 TGGCTGAAGAGCGCCAGATATAGACATAGCCCGCGCGGTGCGCTTACAGCAGGACT 740
Db      602 TGTGTGCGCCACAAATCTGTGATCCACGCGCATGTTTCAAGCTTTTACGAGAGGTGT 661
Qy      741 TTGCGCCCTGCAAAAGGCGCAAGTGGCATCTGCTCAACGCGGCACTATAGAGCCCT 800
Db      662 TCCAGCGCAGAGAAAGAGAAACAATTGAGATCACTCCTCAATGCACTGATCGGAACCTT 721
Qy      801 GGGACAGCAATGAGCTCGGAGCAGAGAGGCTGTGAGCGAGAGATGAATTTTCACTTG 860
Db      722 GGGATGAGACGATTCGCGGAGCAAGAGAGCAGAAACGAGGCGAGGAAATTCAGAGTTCG 781

```

QY	861	CTGTGTTTCCAAATTCATCTTCTTTGAAGAAGCTATCCAGAGAGCATGAAGAAGCAGC	920
Db	782	CTTGGTTTCCTGGAGCCCTTATATACAAGACAGGCGCATACCGAGCTCAATCGGGCTTCAAC	841
QY	921	TGGGCGAGAGGCTTCCAGCCCTCACTCCCGCGGACTTTGCGATCTCAATGCCGGAGAGA	980
Db	842	TCGGGGACCGGTCTACCGGGTTTCACTCCGAGAGGT---CAAAAGTGTACTAGGAAGTT	898
QY	981	CCGACTTCTAGGCGATGATTAATTAACAACATCCAGTTGCGCGCCACCTAGACGTTCCGG	1040
Db	899	CAGATTTTTCGATGATGAACCTCGTACACGACCTTTTTCGTGCGACACAGAGATATGCTTC	958
QY	1041	TCGCCGAGACGGAATATCTCGGCGCCCATTCATGSCATGACACAGAGAAATPAAGCGGACGC	1100
Db	959	CAGATATCAATGACACCAAGAGGCATGTCAATGTGTTACGATACAAACAGCAAGAGGCTCT	1018
QY	1101	CCGTTGGCGAGGAAGAGGCGCTCGCTCGTGGCTCTCTGCCCGGACATGTTCCGGAAGC	1160
Db	1019	CTCGAGGGGAAGATCCGATACGCTTCGTGGCTGGCGSAGGAGCCACCATGATGTGCGCAAGC	1078
QY	1161	ATCTGCGCCGGGGTGTACGGCTGTGACGGCAAGCCCATCTACATCACCGAAGAGGATGGC	1220
Db	1079	TGCTCAATTTGGAATTTGGAACCGAATATATGTGCTTATATATGTTCACGGAATAATG---GCA	1135
QY	1221	CTGTCCCTCGAGAGAGAAACATGACGTGCGAGAGAGGCCGTCAACGACCCCTTCGCGATCC	1280
Db	1136	CGACAGCAAAAAGGAGAGACAGCTCCCAACACAGAGGTACATCTGCACACTTTCCTGATGCG	1195
QY	1281	GSATCTTTGAATCGCATCTGGATCGAATTTCCAAAGCGCATTTACCCCGAGACGGCGTGTGG	1340
Db	1196	GATTCCTTCGAAAGCTACGTTGTGTGTGATTTGGCCGTGTCTTCAAAGAGAGATGTAGTGATATA	1255
QY	1341	TCAAAGGGGATCTTTGCGTGGCGGTTGTGCTGCATAAATTGG	1379
Db	1256	TCCGGTCTTACTTTGCTCGGACCTTCAACCAACAATGGG	1294

```

RESULT 13
US-60-360-039-37092
? Sequence 37092, Application US/60360039
? GENERAL INFORMATION:
? APPLICANT: Cao, Yongwei
? APPLICANT: Chen, Xianfeng
? APPLICANT: Goldman, Barry S.
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
? FILE REFERENCE: 38-10(52052)A
? CURRENT APPLICATION NUMBER: US/60/360,039
? CURRENT FILING DATE: 2002-02-21
? NUMBER OF SEQ ID NOS: 47374
? SEQ ID NO 37092
? LENGTH: 1470
? TYPE: DNA
? ORGANISM: Aspergillus nidulans
? US-60-360-039-37092

```

Query Match	16.0%	Score 317.8	DB 93	Length 1470
Best Local Similarity	53.8%	Pred. No. 3e-57		
Matches 699	Conservative	1	Mismatches 593	Indels 6
				Gaps 2

QY 81 TGCCCAAGCATTTGATATGGGGCTTGCAGAGCGCGCTTACCAGATCGAAGGCGCCGTCA 140  
Db 2 TGCCCAATGACCTTCTTCATGATATGCGACGGCGCGAGGTTGAAGGAGCGTGA 61  
QY 141 AAGAAGGTGCGCGCGCCGTCATCTTGAGACACGTATCGCCACTGGAGCATCGCGCA 200  
Db 62 ACAAGATGGCAAGGGTCTTGTGATCTTGGGACACCTTTGGGCACTACTCAGGCAAAAGTCA 121  
QY 201 CCAACGGCGCCAAAGCGCATGTGGCTTGCGATCACTTACACCGCTTACAGATGAGGACTTTG 260

Db	122	AAGACAATATGCAATGCGATGACCGCTGTGAGGTTCTACACTTTTAAACCGCAAGATGTTG	181
QY	261	ATCTCTTGACCAAGTACGGCGCAAGGCGCTACCGGCTCTCTTGTGTTGTTGCGCGATCA	320
Db	182	CCCTGATGAAGTCGATGATGTGTCACACGGCTTATTCGATTTCTCGCTATCTGTGTGCGCATTA	241
QY	321	TTCCCTCGGGGGGAGGCGTGGATCCCGTTCACACGAGAGGGAATTGATTTTACAGCAAC	380
Db	242	TCCCACTGGAGAGCGCCGACGACCGCGGTCAATGACCAAGGATCAAGTACTACAGGACC	301
QY	381	TGATTTAGCGCCCTGTTTGAAGCGGGGTATCACGCGCTTGGGTGACTTTTGAACCACTGGATC	440
Db	302	TGCTAGACGAGTACTCAATATATGGGATTTAGCGCAATGTGTAACCTCTTTGACTGGGANTG	361
QY	441	TGCTCTAGAGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTTGGAAGAGTTCAGCTG	500
Db	362	TTCCGCAAGCTTTGGAGATCGGATCGGGGGATCTCAACCGAGAGAGGTTCAATTCCTG	421
QY	501	ACTTTGACCGGATATCGAGTTGTGCTTTGAACGTTTGGGAGCCGAGTCCGAACTGGA	560
Db	422	ACTTTGTCCGCTACCGCGGTGTCTGTTTGAACGCTTACGGCCAAAGTCCGTCACTGGA	481
QY	561	TCACCATTACGAGCCCTGATTCAGGCTTACAGCCATCTATGATATGCAACCGGACGAACGCC	620
Db	482	TTACATTTCAATGAACCTGTGTGTACTGCTGGCAGGTTATGCGCAGAGTGTTCATGCTC	541
QY	621	CGGGCAGAGGACGACTTAAACAAGCACTCCACCGAGGCAACACTGCCCTAGCCGCTGGC	680
Db	542	CTGCACGGTCACTCGTTTCCGCACTCAATGAAGAGGATGATTTCTGCAACAGAACCTGTCCA	601
QY	681	TCGCTGAAAAGGCCCAATCATAGACCATGCCCCGCGCGTGGGCGCTTACACAGCAGGACT	740
Db	602	TCGTGGGCAACAAAACTGGTGTGCCACCGGCAATGTTTCCAAAGTTTACCGAGAGTGT	661
QY	741	TTTCGCCCTCGCAAAAAGGCCCAATTCGGCACTCTGGCTCAAGGCGACTATATGAGCCCT	800
Db	662	TCCAACGCGACAGAAAGAACCAATTTGGCATCACCTTCATAGGCAACTGTGTGAAACCTT	721
QY	801	GGGACAGCAATAGAGCTTCGGGACAAGAGAGCTGTGACGACGACGATGGAATTTCAATTG	860
Db	722	GGGATGAAGAAGATCCGGGGGACCGAGAGACGACGAAGCGGCGCAGGGAATTCCAGATCG	781
QY	861	GCTGGTTTGCCAATCCCATCTTTCTTGAAGAAGACTTCCAGAGACATGAAGAAAGCAC	920
Db	782	CTTGGTTCGCCGAGCCCTTATATACAAGACAGCGACTACCCAGCCTCAATGGCGGCTCAAC	841
QY	921	TGGCGAGAGGCTTTCAGGCGCTCACTCCCGGGGACTTTGGCATCCTCAATGCCGAGAGA	980
Db	842	TCGGGGAGCCGTACCGGTTTCACTCCGAGGAGT---CAAAACTGTTACTAGGAAGTT	898
QY	981	CCGACTTCTACGCGATGAATTTACTACACATCCAGTTTCGCGCGCACCTAGACGGTCCG	1040
Db	899	CAGAAATTTTACGATATGAACCTGTAACAGCACTTTTTCGTACACACACAGGATACGCTC	958
QY	1041	TCCCCGAGACGGAATTATCTGGCGGCCCATCCATGAGCAACAGABAAATATAGACGCGACC	1100
Db	959	CAGATATCATATGACCAAAAGGCAATGTCTATTTTTCACGATACAAACGCAAAAGCGTCT	1018
QY	1101	CCGTTGGCGAGAGAGCGGCTCGCTGTGCTGCGCTCTCGCCGAGCAATGTTCCGAGAC	1160
Db	1019	CTCGAGGCGAAGATCCGATACGCTTGTGGGTGGGAGCGCACCACTGGAATGGGGCAAGC	1078
QY	1161	ATCTTCGCCCGGGTGTACGGCTGTATGGGCAAGCCCATCTTACATACCGAGAACGATGCC	1220
Db	1079	TGCTCAATTTGGATTTTGAACCCGATATATGCTGCTTATATATGTCAACGGAAATATG---GCA	1135
QY	1221	CGTCCCTCGAGAGAGAAACATGACGTGCGAGAGAGGCGCTCAAGCAAGCCCTTCGCAATCC	1280
Db	1136	CGAAGAGAAAAGAGAGACAGCTCCCAACCAAGAGTATCTATGACACTTTCCGTATGC	1195
QY	1281	GSTACTTTGACTCGCACTTGACTCGATTTTCAAAGGCCATTTACCAGAGCGCGTCTGCG	1340
Db	1196	GATTTCTTGAAGGCTACGTTGTTGAGATTTGGCCCGTGTCTTCAAGAGAGATGGAATGATA	1255

QY 1341 TCAAGGGTACTTTCCTGGGCGTTCGATAACTGG 1379  
Db 1256 TCCGGTCTTACTTTCCTGGGCGTTCGATAACTGG 1294

RESULT 14  
US-60-138-103-7477  
; Sequence 7477, Application US/60138103  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Timberlake, William E.  
; TITLE OF INVENTION: ASPERGILLUS NIDULANS GENOME SEQUENCE AND USES THEREOF  
; FILE REFERENCE: 38-10(15485)D  
; CURRENT APPLICATION NUMBER: US/60/138,103  
; NUMBER OF SEQ ID NOS: 28006  
; SEQ ID NO 7477  
; LENGTH: 2914  
; TYPE: DNA  
; ORGANISM: Aspergillus nidulans  
US-60-138-103-7477

Query Match 15.4%; Score 307.6; DB 69; Length 2914;  
Best Local Similarity 53.8%; Pred. No. 5,4e-55;  
Matches 701; Conservative 1; Mismatches 595; Indels 7; Gaps 3;

QY 77 GCTCTGCCCCACGACTTTGATGAGGGCTTGGCAAGCGCCCTTACCGATGCAAGGCGCC 136  
Db 1147 GCCTTGGCAGATATGACTTTCATGATATGCGCAGCGCGCGCGGCTTGAAGGAGCG 1206

QY 137 GTCAAGAAAGTGGCGCGCGCGCGCGCTTCATCTGGGACAGTACTGACCTGAGGCGATCG 196  
Db 1207 TGGAAACAAGATGGCAAGGCGCTTCGATCTGGGACAGTACTGAGCGATCTCCAGGCAAA 1266

QY 197 CGCAACCAAGCGCGCAAGCGCGATGTCGTCGATCACTACCAACCGCTACGATGAGGAC 256  
Db 1267 GTCAAGAAAGTATGCAATGCGGATGACGCTGTGAGGTTTACGACTTTTACCGGAGAT 1326

QY 257 TTTGATCTTTGACCAAGTACGCGCGCAAGGCGCTTCTCTTGTGTGTGTCGCGG 316  
Db 1327 GTTGCCCTGATGAAGTGTATGTGTCAACGCTTATGATCTGCTATCTGTGTGCGC 1386

QY 317 ATCATTCCTCCGCGCGCGCGCGCTGATCCGCTCAAGGAGGAAATTGATTTACAGC 376  
Db 1387 ATATATCCACTCGAGGCGCGCGCGCGAGACCCGCTCATATAGCAAGGATCAAGTACTACAG 1446

QY 377 AAATGATTTGACCGCTTGTAGCGCGGATATCAAGCGCTTGGTGAATTTGTAACACTGG 436  
Db 1447 GACTGTGTAGACGACTACTCAATATAGGATTTAGCGCATTCGTACCTTTCACTGG 1506

QY 437 GATTCGCTCAAGCGCTTCAAGATCGCTATGAGGCTGCTCAACGTGGAAGAGTCCAG 496  
Db 1507 GATGTTCGGAAGGCTTTGGAAGATCGGTACGCGGCGCATGTGTAACGAGAGGTTTATT 1566

QY 497 CTGACCTTGAAGGCGTATGAGGCTTGTGCTTGAAGCTTTGGGAGCCAGTCCAGAAC 556  
Db 1567 CTGACTTGTTCGCTACGCGCGCTGTCTGTTTCAAGCGCTTGAAGGCGCAAGGTCGTAC 1626

QY 557 TGGATCAACATCAACGAGMCCCTGG-ATTCAAGCCATCTATGATATGCAACCGGACAGAA 615  
Db 1627 TGGATTTACATTTCAATGAAGATGTGTGTACTGCTGAGGAGTTATCGGAGGTTTCA 1686

QY 616 CGCCCGCGGAGAGCAAGCATTTAACAGACTCCACGAGGCGCAACTGCACTGAGCC 675  
Db 1687 TGCCTCCGACGCGTATGTTTCCGGAATCAATGAAGAGGTTATCTTCGACAGAAC 1746

QY 676 GTGGCTGCTGGAAGGCGCAAGATGAGGCGATGCGCGCGCTGCGCGCTTACAGCAG 735  
Db 1747 GTTCATGTGGCGCACCAAAATGTGTACCCACGCGCGATGTTTCAAGCTTTACAGAGA 1806

QY 736 GGAATTTGCGCCCTCGCAAAAGGCGCAAGATCGGATCTGCTCAACGCGCATCTATGA 795

Db 1807 GGTTGTTCAACCGCAGCAAGAAAGAACAATTGCGATCACCTTCATGCGACTGTGCGA 1866  
QY 796 GCCCTGGAGCAGCAATGAGGCTCTGGAGCAAGAGGCTGTGAGCGACGATGAAATTTCA 855  
Db 1867 ACCTTGGAGTGAAGAGATTCGGGAGCAAGAAAGCAGACCGGCGGCAATTTCCA 1926

QY 856 CATTCGCTGTGTTGCAATTCATCTCTTTGAAGAAAGTATCCAGAGCATGAAGAA 915  
Db 1927 GATCGCTGTGTTCCCGAGACCTTTATCAAGACAGGCACTACCGAGCTCAATGCGGCG 1986

QY 916 GCAGCTGGGCGAGAGGCTTCCAGCCTCACTCCCGGAGACTTTGCCATCTCAATGCCG 975  
Db 1987 TCAACTGGGAGCGGTCTACCGGTTTCACTCCGAGAGT---CAAAATGTTACTAGG 2043

QY 976 AGAGACGACTTATGAGCATGATTAATACATCATCCAGTTGCGCGCCACCTAGAGCG 1035  
Db 2044 AAGTTAGAAATTTTACGATATGAACTGTACAGACTTTTGTGACAGCAAGATAC 2103

QY 1036 TCCCGTCCCGAGACGAGATCTCGCGCGCATTCATGAGCACAGAGAAATTAAGACG 1095  
Db 2104 GCCTCCAGATATCAATGACCAAAAGCAATGTCATTTTCAAGATCAAAAGCAAGAG 2163

QY 1096 CAGCCCGTGTGGGAGAGAGCGGCTCGCTGCTGCGCTCTCCCGGAGCATGTTCCG 1155  
Db 2164 CGTCTCTGAGGCGGAAGAGTCCATACGCTTGTGCTGCGGACGCAACCTGATGCG 2223

QY 1156 GAGCATCTTCGCGCGGCTGTAGCGCTGTACGCGCAAGCCCATATCACTACCGAGAACG 1215  
Db 2224 CAAGCTGCTCAATTTGAAATTTGAAACGATATCAATGCTTATATGTCACGGAATATG- 2282

QY 1216 ATGCGCGTCCCTGAGAGAGAGAAATGACGTGCGAGAGAGCGCGTCAACGACCCCTTCG 1275  
Db 2283 --GCAGAGACGAAAGAGAGAGACAGCTCCCAACAGAGGTACTATGACACTTTCG 2340

QY 1276 CATCGSTACTTGTGATCTGCACTTGAATCTGATTTTCAAGGCGCATTAACAGAGCGCT 1335  
Db 2341 TATGCAATCTCTGAAAGGCTACCTGTGTGATGCGCGCTGTGTCAAGAGATGAGT 2400

QY 1336 CGTGTCAAGGCGTACTTTCGCGGCGTTCGTGATTAATTGG 1379  
Db 2401 GATATTCGCTTACTTGTGCTGACCTTACCGACACTGAG 2444

RESULT 15  
US-09-404-520-19929  
; Sequence 19929, Application US/09404520  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Ghodse, Azita  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: McIntosh, James  
; APPLICANT: Timberlake, William E.  
; APPLICANT: Yu, Jaehyuk  
; TITLE OF INVENTION: Emmericella nidulans Genome Sequence and Uses thereof  
; FILE REFERENCE: 38-10(15498)A  
; CURRENT APPLICATION NUMBER: US/09/404,520  
; NUMBER OF SEQ ID NOS: 44345  
; SEQ ID NO 19929  
; LENGTH: 1434  
; TYPE: DNA  
; ORGANISM: Aspergillus nidulans  
US-09-404-520-19929

Query Match 14.8%; Score 294.6; DB 20; Length 1434;  
Best Local Similarity 53.7%; Pred. No. 2,8e-52;  
Matches 676; Conservative 1; Mismatches 575; Indels 7; Gaps 3;

QY 122 CAGATGAAAGGCGCGCTCAAGAAAGTGTGCGCGCGCGCTTCATCTGAGGACAGTACTGC 181  
Db 3 CAGGTGGAAGACGTGGAACAAAGATGTGCAAGGCTCTTGTGATCTGAGACATTCGCG 62

QY 182 CACTGAGCATTCGCGCAACAGCGCGCAAGGAGATGTGCTTGCATCACTACAC 241

63 CATACTCCAGGCAAAAGTCAAAAGACATATAGCAATAGCGATGCGATGAGGTTCTACAC 122  
242 CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACCGGCAAAAGGCTTACCGCTCC 301  
123 TTTTACCGGCAAGATGTTGCTCTGATGAAAGTGTATGATGATGATGATGATGATGATG 182  
302 TTGTGTTGTTGCGGATCATTTCCCTCGGCGGCAAGGCTGATTCCTGTCACAGAGAGGA 361  
183 CTATCTGTTGCGGCAATATCCACTCGAGGCGGCGACGACCCGGTCAATGAGCAAGG 242  
362 ATTGAGTTTACAGCAAACTGATTAGCGCTGTTGAGCGGGGATTCAGCGCTTGGGTG 421  
243 ATCAAGTACTACAGAGACTGATGACGATCACTCAATTAATGAGATTACCGCATTCGTA 302  
422 ACTTTGTACCACTGGGATCTGCTCAGCGGCTTCACGATCGTATGAGGCTGGCTCAAC 481  
303 ACCCTCTTCACTGGGATGTTCCGAGGCTTTGAGAGATCGGTACGGGGGCAATGCTGAC 362  
482 GTGGAAGAGTTCAGCTGATCTTGAAGCGTATGCGAGTTGTGCTTGAACGTTTGGG 541  
363 CAGGAGAGGTTCAATTCTGACTTGTCCGCTACGGCGGTCTGTTTGAACGCTTAGG 422  
542 GACCGAGTCCAGAACTGATCAACATCAACGACCTTG-ATTGAGGCACTATGATA 600  
423 CCAAGAGTCCGTCACATGATTAATTAATGAAATCATGATGATGATGATGATGATGATG 482  
601 TGCCACCGGCAAGCAACGCGCGGCGAGGAGCACTTAACAAGCACTCCACGAGGCA 660  
483 TGCGGAGAGTGTCTGCTCTGCGAGGTCATCGTTTCCGAACTCAATGAAAGAGTGA 542  
661 CACTGCCACTGAGCCGTTGCTGCTGAAAGCCCAATCATGAGCCATGCGCGCGCT 720  
543 TTCTCGACAGAACCGTTTCAATGTCGCGCACACAAACTGGTGAACCCAGGCAATGTTTC 602  
721 GCGCGTCTACAGCAGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGCACTGCTCA 780  
603 CAAGCTTTACGAGAGGTTTCACACCGCAGAAAGAACAATTGGCATCACCTTCA 662  
781 CGGCACTACTATGAGCCTGAGACAGCAATGAGCCTGAGCAAGAGGCTGCTGAGCG 840  
663 TGGCACTGTCGAGAACTTGGGATGAAAGCATCCGCGGACCAAGAAAGCAGACCG 722  
841 ACGATGGAATTTCACTTGGCTGTTTGCATTCATCTTCTTGAAGAAAGTATCC 900  
723 GGCAGAGGAATTCAGATCGCTTGGTCCGAGCCCTTATACAAGCAGGCGACTACCC 782  
901 AGAAGCATGAAGAGCAGCTGGGCGAGAGGCTTCCAGCCTCACTCCGCGACTTGC 960  
783 AGCTCAATGCGGGCTCAACTCGGGGACGTTTACCGGTTTCACTCCGAGAGT---C 839  
961 CATCTCAATGCGGAGAGACGACTTCTACGGCATGAATTAACAATCCAGTTCCG 1020  
840 AAAACTGTACTAGGAATTCAGAAATTTACGATATGAATCTGTAACGACCTTTTGT 899  
1021 GCGCACTTACAGGTCCTGCTCCGAGACGCACTATCTCGCGCATCAATGACACA 1080  
900 GCAGACAAGGATAGCTTCCAGATATCAATGACCAAAAGCAATGTATGTTACGA 959  
1081 GGAGAAATAAGACGCGACCGCTTGGGAGAGAGCGGCTCGCTGCTGCTGCTG 1140  
960 TACAAACGCAAGAGCGTCTCTCGAGGCGAAGATCCGATAGCCTTGGCTGCGAGCGC 1019  
1141 CCCGACATGTTCCGAGACATCTGCGCGGAGTGAAGGCTGTAAGGCAAGCCATCTA 1200  
1020 ACCCACTGGATGGCGCAAGCTGCTCAATTTGATTTGAAACCGATATCATGTGCTATATA 1079  
1201 CATCACCGAAGACGATGCCGTGCTGAGAGAGAAATGACGTGCGAGAGCGCT 1260  
1080 TGTACGGAATAATG---GCAGACAGCAAAAGAGAGACAGTCCACACCAAGGATCT 1136  
1261 CAAGACCCCTTCCGATCCGCTACTTTGACTCGGACTTGGACTGATTTCCAAAGCCAT 1320

Db 1137 CATGACACTTTCCGATAGGATTTCTGAAGGCTACGTTGATGACCCGCTGT 1196  
Gy 1321 TACCCAGACGCGCTGCTGCTCAAGAGGAGTACTTTGCGTGGCGCTGCTGATAACTTGG 1379  
Db 1197 CAAGAGATGAGATGATATCCGCTTACTTTGCTGACCTTCAACGACAACTGG 1255

Search completed: March 29, 2004, 02:50:59  
Job time : 6866.96 secs



LENGTH: 2016  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-152-372-253

Query Match 9.0%; Score 178.2; DB 6; Length 2016;  
Best Local Similarity 54.5%; Pred. No. 6.9e-36;  
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```
QY 76 AGCTCTGCCCAACGACTTTGTAATGGGGCTTGCACAGCGCCGCTTACAGATCGAAGCGC 135
DB 237 AACCTCCCTCTTGGCTTCTCTGGGGGCTGGGAGTTCTTGCACAGAGGAGGCGC 296
QY 136 CGTAAAGAGAGTGGCCGCGCCGCTTCCATCTGGAGACAGTACCTGACCCAGCCGATC 195
DB 297 CTGGAGCCAGAGCGGAAAGGGCTTAGCATCTGGAGCGTCTTTCACACACAGTGGAGAGG 356
QY 196 GCGACCAACGAG---CGCCAAAGCGGATGTGGCTTGGCATCACTACCAAGGCTACGATGA 252
DB 357 GAAAGTCTTGGAGATGAGACGGCAGATGTAGCCTGTGACGGCTACATCAAGTCCAGGA 416
QY 253 GGAATTGATCTTGTGACCAAGTACGGGCAAGGCTTACCGCTTCTCTTGTGTGTC 312
DB 417 GGAATCATCTCTGCTGAGGAGAACTGACGTCAACCACTACCGATCTCTCTGTGTGCCC 476
QY 313 GCGATCATTTCCCTCGCGCGGAGGCTGGATCCGCTCAAGAGAGGAGGGAATTGATTTA 372
DB 477 CGGGCTCTGCCCACAGGCAATCCAGCCGAGCGAGTGAACAAGAGGGAATCGAATTTTA 536
QY 373 CAGCAAACTGATTAAGCGCCCTGTGAGGCGGGGATATCACGCTTGGTACTTTGTATCA 432
DB 537 CAGTGATCTTATCATGATGCTCTTCTGAGCAGCAATCATCTCCATCTGATCTTGACACA 596
QY 433 CTGGATCTGCTTCAAGCGCTTCAAGATGCTGTAGAGGCTGGCTCAACGTGGAAGAGT 492
DB 597 CTGGATCTGCGACAGCTGCTCAAGTCAAAATACGCTGGTGGCAGAAATGTGAGCATGGC 656
QY 493 CCAAGTGAATCTTGAAGCGGTATGCGAGGTTGCTTGAACGTTTGGGAGCCGAGTCCA 552
DB 657 CAA---CTACTTCAGAGACTAGCCCACTGTGCTTTTGAAGCCTTTGGGAGCCGTTGAA 713
QY 553 GAACTGATCAACATCAAGCAAGWCCCTGTGATTCAGGCACTATGTGATATGCCACGGGAG 612
DB 714 GCACTGATCAAGTTCAAGTATCTTGGGCAATGGCAGAAAGGCTATGAGACGGGCA 773
QY 613 CAAGCCCCCGGAGAGGAGCATTAACAAGCACTCCACGAGGGCAACACTGCCACTGA 672
DB 774 CCATGCGCGGGCTTGAAGCTCCGC-----GGCAGCGG 806
QY 673 GCGGTGGCTCGCTGAAAAGGCCAGATCATGAGCCATGCGCGCGCTGTCTACAG 732
DB 807 CTGTGACAAAGGAGACACACATCATTAAGGCCCAAGCAAACTTGCACTTTTATTA 866
QY 733 CAGGAGCTTTCGCCCTCGCAAAAGGCGCATTCGCAATCTGCTCAACGGCGACTACTA 792
DB 867 CACCAAGTGGGCGACAGCAAGCAAGGCTCTGGTGGAAATTTCACTGAAGTGAAGGAG 926
QY 793 TGAGCCCTGGGACAGCATGAGCTTCGGGACAAAGAGGCTGTGAGCAAGATGGAATT 852
DB 927 GGAACCTGTGACATTAATTAACCCCAAGAGCCTTAAGGCTGCGAGATTAATCAAGT 986
QY 853 TCACATGCTGCTGTTTGGCAATCCATCT 881
DB 987 CTGTCTGGGCTGTGTTGCCAACCCTTT 1015
```

RESULT 2  
PCT-US04-02242-95  
Sequence 95. Application PC/TUS0402242  
GENERAL INFORMATION:  
APPLICANT: Diversa Corporation  
APPLICANT: Barton, Robertson, Dan; Elkins, James; Chang, Kristine  
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN

TITLE OF INVENTION: THEM  
FILE REFERENCE: 5646-2018-40  
CURRENT APPLICATION NUMBER: PCT/US04/02242  
CURRENT FILING DATE: 2004-01-26  
PRIOR APPLICATION NUMBER: 60/442,794  
PRIOR FILING DATE: 2003-01-24  
NUMBER OF SEQ ID NOS: 224  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 95  
LENGTH: 1374  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
PCT-US04-02242-95

Query Match 8.2%; Score 162.8; DB 1; Length 1374;  
Best Local Similarity 60.0%; Pred. No. 5.2e-32;  
Matches 329; Conservative 0; Mismatches 207; Indels 12; Gaps 3;

```
QY 81 TGCCCAACGACTTTGATGGGGCTTGCACAGGCGCCGCTTACAGATCGAAGGCGCTCA 140
DB 38 TTCCCGCACTTTCACCTTGGGGGCTGGCCACAGCGCTTACAGATCGAAGGCGCGCG 97
QY 141 AAGAGGTGCGCGCGCGCTTCATCTGGGACACGTACTCCACCTGGAGCATCGCGCA 200
DB 98 CCATCGGCGCGCGCGCGCTTCATCTGGGATACCTTCACGCAACGGAAGGCAAGATCA 157
QY 201 CCAAGCGCGCAAGGAGATGTGCTTGCATCACTACCAACGCTATAGAGAGACTTTG 260
DB 158 TCGAGGCAACAGTGGGAGCTGCGCTGCGACACACTACCGCTATGCGGAGACTGG 217
QY 261 ATCTTGAACCAAGTACGCGCAAGGCTTACGCTTCTCTTGTCTGTGTCGGATCA 320
DB 218 AGCTGATCGCAGCTTGGGCTGAGAGCGCTTACAGCTTTCATGATCTGTCGCGCTCC 277
QY 321 TTCCCTCGCGGCGAGCTGATCCGCTCAACGAGAGGGAATTTGATTTTACGCAAC 380
DB 278 AGCCACAGGGGTCC-----GCGCTGGAACGAAGAGGCTTGAATTTTATGCCCCG 331
QY 381 TGATTAAGCGCCCTGTGAGGCGGGGATACAGGCTTGGTGAATTTGATCACTGGAGT 440
DB 332 TGTCTGAGCGCTTGGCGGCAAGGACTGACGCGACCTGACCTTATCACTGGAGAC 391
QY 441 TGCTCAGCGCTTACAGATGCTATGAGGCTGACTCAACGTTGAAAGAGTCCAGCTGG 500
DB 392 TGCCGCAAGCTTTCAGAGA---CGAGGCGGCTGTGCTCAATCGCGCACTGTAC--- 445
QY 501 ACTTGAAGGATGACAGGTTGTCTTGAAGCTTTTGGGAGCCAGTCCAGAACTGGA 560
DB 446 ACTTGGCGCGTATGCGCGCGAGGTTGGCGCGCTTTCGACCAAGATGCGCAGCATCG 505
QY 561 TCACATCAAGAACCCCTGATTCAGGCAATCTATGATATGACCGGCGAGCAAGCGCC 620
DB 506 CCAGCAATGAGCGGTGTGACCTGCTGTGGCCAGGACCGGCAAGTTCGCGC 565
QY 621 CCGGCAAG 628
DB 566 CCGGATG 573
```

RESULT 3  
US-10-767-701-7585  
Sequence 7585. Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10767,701  
CURRENT FILING DATE: 2004-01-29



```

; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 7585
; LENGTH: 805
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS93553_1
; US-10-767-701-7585

```

```

Query Match      8.0%; Score 159.8; DB 6; Length 805;
Best Local Similarity 54.3%; Pred. No. 2,4e-31;
Matches 345; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

```

```

QY 81 TGGCCAAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTACGATGGAAGCGCCCTCA 140
DB 174 TCCCAAGAGACTTGTGTTGGGACAGAGCTCCGGGGTATCAATAGAGGTGCTTACA 233
QY 141 AAGAAGTGGCCGGCCCGCTCCATCTGGGACAGTACCTGCACTGGAGCCATCGCGCA 200
DB 234 ACGAAGAGGCAAGGGGCTTACCATATGGGACAAGTTCACTACATCCAGTAAATCT 293
QY 201 CCAAGCGGCAAGCGGATGGCTTGGCATCACTACCGCTAGATGAGACTTTG 260
DB 294 TGAACAACGATACCGGAGAGTACAGATGACATGATCAATCAAGAGAGTGTG 353
QY 261 ATCTTGAACCAAGTACGGGCAAGGCTTACCGCTTCTCTTGTCTGGTTCGGGATCA 320
DB 354 AACTCTCAAGAGACATGAACTTGGAGCCCTTCGGCTTCTCATTTGCTGACAGATCC 413
QY 321 TTCCTCTGGCGGAGGCTGATCCGTCACAGAGGGAATTTGATTTTACAGCAAC 380
DB 414 TGGCAACGAGTCCCTGATGAGAGAAATCAACAAAGAGCGTGGCTTCTATTAACAA 473
QY 381 TGAATGACGCTTGTGAGGCGGGGTATCAAGCTTGGGTGACTTTTACACTGGGATC 440
DB 474 TCAATCAACGAGTATGAGCCAAAGGCTGAAGCCATTGTACCATTTTACACTGGGACA 533
QY 441 TGCCTCAGGCGCTTACGATCGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAGCTG 500
DB 534 CGCCCTCGGCTTGGAGGACAGATGAGATTTCTCAGCGAAGACATCATCA---GG 590
QY 501 ACTTTGAGCGTATGCGAGGTGTGCTTTGAAACGTTTGGGAGCCGAGTCCAGAACTGA 560
DB 591 ACTACGAGGACTTGCAGAGTTTGTCTTCAAGAGTTTGGCGAGCCGCTCAAGGCTTGA 650
QY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACGGGACAGACGCC 620
DB 651 CCAAGTTCAACGAGCCGTGACGTAATGCAAGGCTACGCGCTGCGCAAGTCCGCGC 710
QY 621 CGGGCAGGACAGCATTAACAAGACTCCACCGAGGCAACACTGCACTGAGCCGTGGC 680
DB 711 CGGGCGGCTGTGTGTGATGATCAACAAAGATGCTTCCCGGCGAGCTCGGCGGAGC 770
QY 681 TCGCTGAAAAGGCCAGATCATGAGCCATGCGCGC 715
DB 771 CTAACACCGTGAAGCAACATCATCTCGCCAC 805

```

```

RESULT 4
US-10-767-471-414
; Sequence 414, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 414
; LENGTH: 6274
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(6274)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-5)
; US-10-767-471-414

```

```

Query Match      7.6%; Score 151.6; DB 6; Length 6274;
Best Local Similarity 52.4%; Pred. No. 7,4e-29;
Matches 446; Conservative 1; Mismatches 371; Indels 33; Gaps 4;

```

```

QY 81 TGGCCAAAGCACTTGAATGGGGCTTCGCAACGGCCGCTTACGATGGAAGCGCCCTCA 140
DB 4141 TTCCTAAGGCTTCAATGTGAGAGTGCAGCTTCTCTCAATACATGATGGAAGGCTTGA 4200
QY 141 AAGAAGTGGCCGGCCCGCTCCATCTGGGACAGTACCTGCACTGGAGCCATCGCGCA 200
DB 4201 GAGCAATGCAAGAGACTTCAGCATTTGGGACAGTTTCTCACACCACTGAGGTTG 4260
QY 201 CCAAGCGGCAAGCGGATGGCTTGGCATCACTACCGCTAGATGAGACTTTG 260
DB 4261 AGAAGATGCAATGGAAGAGTGGCTCTGACAGTTATCAAGATTTGCTGAGGATCTGG 4320
QY 261 ATCTTGAACCAAGTACGGGCAAGGCTTACCGCTTCTCTTGTCTGGTTCGGGATCA 320
DB 4321 TCACCTGCAAGACTGGGGTGTCCACTACCGTTTTCATCTCTGCTGCTCGCATCC 4380
QY 321 TTCCTCTGGCGGAGGCTGATCCGTCACAGAGGGAATTTGATTTTACAGCAAC 380
DB 4381 TCCCTGATGACACCAAGGTA---CATCAATAAAGCGGCTGAACTACTAGTGGGCT 4437
QY 381 TGAATGACGCTTGTGAGGCGGGGTATCAAGCTTGGGTGACTTTTACACTGGGATC 440
DB 4438 TCAATGATACACTGTGCGCCGACGATCCAGGCCAGAGTACATTTACACTGGGAGC 4497
QY 441 TGCCTCAGGCGCTTACGATCGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAGCTG 500
DB 4498 TACCAAGAGCGCTCAAGAT---GTAGAGGCTGGGAGAAATGAGACATGTGTACAGCG 4553
QY 501 ACTTTGAGCGTATGAGAGGTGTGCTTTGAAACGTTTGGGAGCCGAGTCCAGAACTGA 560
DB 4554 --TTTAAAGAGTATGAGATGTGCTTCCAGAGGCTGGGAGCAAGATGAAATTTTGA 4611
QY 561 TCACATCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACGGGACAGACGCC 620
DB 4612 TCACGTTGAATGAGCCCTTTGTATGCTTACAGAGGCTATGCTACGAAACAGAGCTC 4671
QY 621 CGGGCAGGACAGCATTAACAAGACTCCACCGAGGCAACACTGCACTGAGCCGTGGC 680
DB 4672 CAGGAGTCTCA-----ATAGGCTTGGCACTGCCCTTACA 4707
QY 681 TCGCTGAAAAGGCCAGATCATGAGCCATGCGCGCCGCTGTACAGAGGAGCT 740
DB 4708 TTTTGGCCCACAATCTAATAAAGCTCATGAGGCTGAGCTGTGTCMAAGATGTGT 4767
QY 741 TTCGCCCTTGGCAAAAGGCGCAGATCGCATCTCGCTCAACGGCGACTATAGAGCCCT 800
DB 4768 ACCGCGCACTCAAGGTGGCTGATTTTCATCACTACAGATGATGGGCTGAACCA 4827
QY 801 GGAAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCGAGCATGGAATTTTCAATTG 860
DB 4828 GAGATCCCTTCAACGAGAGATGTGAGGACCGAGAGATATGTTGATGATGAGGAG 4887
QY 861 GCTGTTTGGCAATCCCATCTTTTGAAGAAGACTATCAAGAGACATGAAGAAGCAGC 920
DB 4888 GCTGTTTGGCACTCTATTTTCAAGAAATGAGATTAACANTGAGTATGAAGAAGCAGC 947
QY 921 TGGCGAGAGG 931
DB 4948 TCCGTGACAGG 4958

```

RESULT 5

```

US-10-767-701-6851
; Sequence 6851, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5353) B
; CURRENT APPLICATION NUMBER: US/10/767,701
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6851
; LENGTH: 767
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(767)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS5746_1
US-10-767-701-6851

```

```

Query Match
Best Local Similarity 7.4%; Score 147.8; DB 6; Length 767;
Matches 352; Conservative 0; Mismatches 269; Indels 14; Gaps 3;

QY 77 GCTTGCCCAACGACTTGAATGGGGCTTCGCAACGGCGCCCTACAGATCGAAGCGCC 136
DB 147 GCGTTCGCGAGGAGGCTTCGTCGCGGAGCGCGCTCGGCTACAGATCGAAGCGGATG 206
QY 137 GTCGAAGAAGGTGGCGCGCGCGCTGCATCTGGGACAGTACTGCCACCTGGAGCCATCG 196
DB 207 GCCAAGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
QY 197 CGACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 256
DB 267 ATCCCTAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
QY 257 TTTGATCTCTTGAACGAAGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 316
DB 327 GTGAACATATGAGAACATGAGGCTTTGATGCGTACCGGTTTCAATTTCTTGATCGAG 386
QY 317 ATCATTTCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376
DB 387 ATTTTCCCAATGAACTGGCAAG-----GTGAACGAGAAAGAGATGATTAATTAAC 440
QY 377 AAACGATGAGCGCGCTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436
DB 441 AGGCTCATAGATTATGATGCTTTCAGCAAGGATACGCGCGTATGCAAAATCTTACATTA 500
QY 437 GATTCGCTCAGCGCGCTTACGATGCTATGAGGCTGCTCAACGATGAGGAGGTCAG 496
DB 501 GACCTCCCATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 557
QY 497 CTGACCTTGAAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCGAAC 556
DB 558 GAGGCTTTGACAGTACGCGCGAGTGTGCTTTCAGACGTTCCGAGACGAGGTTGAAGAC 617
QY 557 TGAATCAGCATCAAGMCCCTGATTCAGGCGCATGATGATGATGATGATGATGATGAT 616
DB 618 TGGTTTACTTCAACGAGCGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 677
QY 617 GCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 676
DB 678 GCANCGG-----GAGGTGTTCGAGTGGCGCGCGGAGGCAATTCAGACGAGAGCG 732
QY 677 TGGCTGCTGAGAAAGCGCGCATGATGAGGCGATG 711
DB 733 TACCTNGCTGACACCATCTCCTTTCATGC 767

```

```

RESULT 6
PCT-US04-02242-93
; Sequence 93, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANT
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; PRIOR FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-93

```

```

Query Match
Best Local Similarity 7.0%; Score 139; DB 1; Length 1389;
Matches 316; Conservative 0; Mismatches 225; Indels 12; Gaps 3;

QY 72 CGCTAGCTGCGCCCAACGACTTGAATGGGGCTTCGCAACGGCGCCCTACAGATCGAAG 131
DB 26 CCTGTGCTTCCAGAGAGGCTTCGTCGCGGAGTGGCGCGCGCGCGCGCGCGCGCGCG 85
QY 132 GCGCGCTTCAAGAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 191
DB 86 GCGCGCTTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 145
QY 192 CATGCGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 251
DB 146 GAGCGGTTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 205
QY 252 AGGACTTGTATCTTGAACGAAGTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 311
DB 206 AGGAGTGTGCTTGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 285
QY 312 CGCGGATCATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
DB 266 CCGAGTCTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 319
QY 372 ACAGCAACTGATGAGCGCGCTTGTGAGCGGAGTATCAGCGCTTGGGATGATTTGAC 431
DB 320 ACTGCGGTGTGAGACGCGCTGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 379
QY 432 ACTGGAATCTGCTCAGCGCGCTTCAAGATGCTATGAGGAGGAGGAGGAGGAGGAGGAG 491
DB 380 ATTGGAGTACCTCTTGCTCT---CTATACCGGAGGAGGAGGAGGAGGAGGAGGAGG 436
QY 492 TCCGCTGACTTTGAGCGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551
DB 437 CGGATTGG---TTTGGCGAGTACGCGGCGCTATGCGCGATGCGCTCTCCAGCGGATG 493
QY 552 AGAATGATCACATCAACGAGMCCCTGATTCAGGCGCATGATGATGATGATGATGATG 611
DB 494 AGATTTCTTCACTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 677
QY 612 GCAACGCGCGCGG 624
DB 554 AGCATGCTCAGG 566

```

```

RESULT 7
US-10-767-701-14696
; Sequence 14696, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.

```

APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 14696  
LENGTH: 2042  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS4596\_1  
US-10-767-701-14696

Query Match 6.6%; Score 131; DB 6; Length 2042;  
Best Local Similarity 51.5%; Pred. No. 8.5e-24;  
Matches 433; Conservative 1; Mismatches 386; Indels 21; Gaps 5;  
QY 81 TGGCCAAACGACTTGAATGGGGCTTGGCAACGGCCGCTTCCAGATGAGAGCGCCCTCA 140  
DB 165 TCCCGCGGGGCTTCTTGGCGTGGCTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGG 224  
QY 141 AAGAAGTGGCGCGCGCGCGCTCCATCTGGGACAGTACTGCCACCTGAGCCATCGCGCA 200  
DB 225 CAGAGATGGAAGGAAGGAGCTAGCATCTGGGACACATTCACACATTAAGATATTCT--A 281  
QY 201 CCAACGGCGCAACGGGATGTGGCTTGGCATCAACAACCGTACGATGAGGACTTTG 260  
DB 282 TTGACGGTGGCAAGGATGATGTAATCGGATCAGTATCAATGAATCAAGGAAGATGTA 341  
QY 261 ATTCTTGAACCAATGACGGGCAAGGCGTACCGCTTCTCTCTCTCTCTCTCTCTCTCT 320  
DB 342 AGCTTTTGATGAGTGGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 401  
QY 321 TTCCCTCGCGGCGAGGCTGGATCCGTCACAGAGAGAGGAATTTGATTTCAGCAAC 380  
DB 402 TTCTCTGATGCG-----CGAGAGCTGTCAATCCGAAGGGGCTGGAGTACTACACATC 455  
QY 381 TGAATGACGCGCTTGTGAGCGGGGATTCACGCGCTTGGTGAATTTTACCACTGGATC 440  
DB 456 TGAATGATGAACTCTTGAAGTACGGAATACCACTGTAAGATCTATCATTTGATTT 515  
QY 441 TGGCTTACGGGCTTTCAGATCCGCTATGAGAGGCTGGCTCAACGTTGAGAGAGTCCAGTGG 500  
DB 516 TTCTTCAAGGCTTTCAGATGAGTACAAAGGGGCTGTTAGTCTTGA--TTGATAGAGG 572  
QY 501 ACTTTGAGCGTATGAGAGGTTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAACTGA 560  
DB 573 ACTACACAGATACGAGATGTGTCTTCAAGAACTTTGGCGAGAGTGAAGTACTGGA 632  
QY 561 TCACCATCAACGACCTGTGATTCAGGCATCTATGATATGCCACCGGAGCAACGCC 620  
DB 633 GCACTGTCATGAACTTATCATGAAACGATTTGGCGGATTTGACCAAGAACTCTACAC 692  
QY 621 C-----GGGAGAGAGAGATTAACAAGACATCCACGAGGGTAAACTGCCACTAGC 674  
DB 693 CACAGCGATCTCAGCCCTTGGCTTGGCTTGGTGTGATGAAGGCAATCCACCAACAAC 752  
QY 675 CGTGGCTCGGTGAAAGGCGCAATCATGAGCCATGCGCGCGCGTGGCGCTTCAAGCA 734  
DB 753 CTAACATAGACACACACCTTGTACTTGGCGATGATCTGCAGTGTCTCTATATAGAG 812  
QY 735 GGAAGCTTGGCCCTTGGCAAAAGGCGCAATCGGCATCTGCTCAACGGCGCATTAAG 794  
DB 813 AGAAGTACAGGCTGAGCAAGAGAGCAATTTGACAGAGCTTGGTGGTGGTGGTGGTGG 872  
QY 795 AGCCCTGGAGACAGATGAGCTTGGGACAAAGAGGCTGCTGAGCGACGATGAAATTTG 854  
DB 873 AGCTT--GCAACGGAAGACTTCCGATGATATTTGAGAGAGCTGCAAGGATGAAATTTTC 929  
QY 855 ACATTGGCTGTTTGGCAATCCCATCTTTTGAAGAGAGATATCCAGAGAGCATGAAGA 914

DB 930 ACCTGAGTGGTTCATGATCCTATGATGATGAGAGTACCTCCAGTATGAGAGAGA 989  
QY 915 A 915  
DB 990 A 990

RESULT 8  
US-10-093-037A-2  
Sequence 2, Application US/10093037A  
GENERAL INFORMATION:  
APPLICANT: Jay M. Short  
APPLICANT: Bylina, Edward  
APPLICANT: Swanson, Ronald V.  
APPLICANT: Mathur, Eric J.  
APPLICANT: Lam, David B.  
TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF  
FILE REFERENCE: 564462001402  
CURRENT APPLICATION NUMBER: US/10/093,037A  
CURRENT FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: US 09/910,579  
PRIOR FILING DATE: 2001-07-20  
PRIOR APPLICATION NUMBER: US 09/134,078  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 08/949,026  
PRIOR FILING DATE: 1997-10-10  
PRIOR APPLICATION NUMBER: US 60/056,916  
PRIOR FILING DATE: 1996-12-06  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1317  
TYPE: DNA  
ORGANISM: Thermotoga sp.  
US-10-093-037A-2

Query Match 6.6%; Score 130.8; DB 6; Length 1317;  
Best Local Similarity 56.4%; Pred. No. 7.9e-24;  
Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;  
QY 77 GCTTGGCCCAACGACTTGTGATGAGGCTTGCACGCGCGCTTACAGATGAGAGGCC 136  
DB 16 GATTTTCCAAAAGATTTTATCTTGGACGGCTTACCGAGCATACAGATGGAAGTGCA 75  
QY 137 GTCAAGAGAGTGGCGCGCGCGCTCATCTGGGACAGTACTGCCACTGGAGCATCG 196  
DB 76 GCAAGAGAGATGCGAGAGGCGCATCATTTGGGATGTCTTTTACACACGCGTGGCAA 135  
QY 197 CGCAGCAAGCGCGCGCGAGTGTGGCTTGGATCATACACCGCTAGAGAGAGAG 256  
DB 136 ACCCTGAGCGGTACACAGAGAGCTTGGCTGTGACCATTTATACAGATCAAGAGAGAT 195  
QY 257 TTTGATCTTGTGACCAAGTACGCGCGCAAGGCGCTTCTCTGTGTGTGCGG 316  
DB 196 ATCCAGGTATGAAGAAATAGGTTTACAGCTTACAGTTTCTCTCTGCGCCAGA 255  
QY 317 ATCATTCCTCGCGCGCGAGCTGATCCGCTCAAGAGAGAGAAATGAGTTTACAGC 376  
DB 256 ATTATGCGAGATG-----GAAGAACATCAACCAAAAGGCTGTGATTTCTACAC 306  
QY 377 AAACATGACGCGCTTGTGAGCGGGGTATCAGCGCTTGGGATTTGTACACTGG 436  
DB 307 AGACTGTTATAGCTTTTGAAGAAATGATATATACATCATCTTATACACTGG 366  
QY 437 GATTCGCTCAGCGCTTCAAGTCCGTATGAGGCTGCTCAACGTGAAGAGTCCAG 496  
DB 367 GACTTACCTTACGACATTTATGA---AAGGTGATGGTTAACCC---AGATATAGCG 420  
QY 497 CTGAGCTTGAAGGAGTATGAGAGTGTGCTTTGAAGCTTTGGGAGCGAGTCCAGAAC 556  
DB 421 CTCTATTTCAGACATACGCAAGCTTATGTTCAACGAACCTGCGTATGTGTGAACAT 480

QY	81	TGCCCAAGCACTTTAAATGGGGCTTCGCAACGGCGCTACCAAGATCGAAGGCGCCTGA	140
Db	590	TCCCCGAAGGCTTCGTTCCGCGTCGCCACTCCGCGTACCAAGATTTGGGGAGCCAGAA	649
QY	141	AAGAAGTGGCCGCGGCCCGTCCATCTGGGACACGTACTGCCACCTGGAGCCATCGCGCA	200
Db	650	GGGAGGGAGGCAAAAGAGACAGCATATGGGATGTAATTTCACATGACAAAGAAGCGTCT	709
QY	201	CCAAACGGGCCCAAGCGCGATGTGGCTTGGCATCACTACCAACGGCTACGATGAGACTTTG	260
Db	710	TAGACAAAAGCAATGCAGAAATTTGCATGTGACACTACCATCATGATPACAAAGAAACATTG	769
QY	261	ATTCCTTGACCAAGTACGGCGCAAGGCGCTACCGCTTCCTTCGTCGTGGCGGATCA	320
Db	770	AGCTCATGGCAGTGTAGGTTTTCAGCGGTACAGATTTTCTATATCTTGGGACCGTATAT	829
QY	321	TTCCCTTCGGCGGCGCAGGCTGATCCCGTCAACGAGAGAGGAATTAAGTTTACAGCAAC	380
Db	830	TTCC--TGATGGCTTGGGGGAAAAAGTCAATGACACAGAGATCGCCTTCTACAAATGACC	886
QY	381	TGATTGACGCCCTGTTGAGCGGGGTATCAACGCTTGGGTACTTTGTAACACTGGGATC	440
Db	887	TCATCAACTTTATGATTTTCGAAGGATTTGAGCTTACGCAACTCGTATCATTTGGGATC	946
QY	441	TGCGTCAGGGCGCTACGATCGCTATGAGAGTGGCTCAACGTGGAAGAGTTCAGCTG	500
Db	947	TTCCAAACAATCTTCAAAAAAACTCTGGGGGGCTGGATTTCT--GACAAATTTGGAGT	1003
QY	501	ACTTTGAGCGGTATGCGAGTGTGCTTTGAAACGTTTTGGGAGCCGAGTCCGAATCTGGA	560
Db	1004	ACTTTGCATTGTATGACAGAGGTGCTTCTTGCAATTTTGGAGACAGAGTAAGCGTTGGA	1063
QY	561	TCACCATCAACGAWCCCTGGATTCAGGCACTATGATATGATGCCACCGGACAGCAACGCC	620
Db	1064	TAAACATCAATGAGGCTCTCCAAACTGCAATCAATGATGTTATGTAATTTGGAC	1122
QY	621	CGGCGACGAGC	631
Db	1124	CTGGTGTATGC	1134

OY	8	ACGACCTTGAATGGGGCTTGGCAAAAGGCGCCCTACACAGATCGAAGGCGCCGTCAAAAGG	146
Db	33535	ATGACTTTCTGTGGGGGGCTCTCTTCCGCTTATACGATTTGAAGGGCGCTGGGAATGCCG	3359
OY	147	GTGGCCCGCGCCCGCTCCATCTGGGACACGTACTGCCA---CTGAGCCATCGCGACCA	203
Db	33595	ATGGCAAAAGGCCAGCATCTGGGATTAATTACCCACACACGAGGAGCAATGTGAAG	3365
OY	204	ACGGCGCCCAACGGCGATGTGGCTTTGGGATACTAACACCGCTACGATGAGGACTTTGATC	263
Db	33655	ACAAATGCACATGGAGACATGCGCTGTGACAGCTATACCAAGCTGATGCGGACTTAATA	3371
OY	264	TCTTGAACCAAGTACGCGCGCAAAAGCCCTACCGCTTCTCCTGTGCTGCGCGCATCATTC	323
Db	33715	TGCTCCGAGCTTTGAAGGTAAAGGCTTACCGCTTCTCTATCTCTGTCTCGG---ATTT	3377
OY	324	CCCTCGCGCGAGCTGGATCCCGTCAACGAGGAGGAAATGTGATTTCACGAAACTGA	383
Db	33772	TCCCAACTGGGAAAGAACGCTCTATCAACAGCATGATGGGGTGTATTTCACACAGCTGA	3383
OY	384	TTGACGCGCCGTGTGAGGCGGGGATACACGCTTGGGTAATTGTACCACTGGGATCTCG	443
Db	33832	TCATGTGCTTGTGGCAGACCAATCTTCCATGGTGAATGTTCATATGGAGACTCG	3389
OY	444	CTCAGGCGCTTCAACGATCGCTATGAGAGGCTGAGCTCAACGTGAAGAGTCCAGCTGACT	503
Db	33892	CCAGGCGCTTCCAGGAATATC---GGAGGCTGGGAAATCTCGCTTGATTGACTTG---T	3394
OY	504	TTGAGCGGTATGCGAGGTGTGCTTTGGAACGTTTGGGACCGAGTCCAGACTGATCA	563
Db	33946	TTGACAGCTTACGACGACTTTTGTTCACAGCTTTGTGTATGATGACGAACTTTTGGATGA	3400
OY	564	CCATCAACGAMCCCTGGATTCAGAGCATGTATGATATGCAACCGGACAGAACGCCCGG	623
Db	34006	CTTTTATGTAGCCCATGTACTGTGCAATGGCTAGGTTATGTCTTAGGGGAAATTTCCCCAG	3406
OY	624	GCAGGAGCAGACTTAAACAGCACTCCACCGAGGGCAACTGCGCATGAGCCGTGACTCG	683
Db	34066	G-----GATGAAGACCCAGGCTGGGACCATATATAGA	3409
OY	684	CTGGAAGAGCCGAGATCTAGAGCCATGCCCCGCGCGCTGACCGCTTACACAGAGGACTTTC	743
Db	34099	TACCCACGCGCTCATTAAGACCCCAATCCAGAACTTATCAACGTAACGATGAGAAATACA	3415
OY	744	GCCCTCTGCAAAAGGGCCGAGATGGGATCTGCTCAAGGGCACTACTATGAGCCCTGGG	803

Db 34159 GGCAGGAGCAGAAAGGGGCTCTCTGCTGAGCCCTCAGTACACACTGGGACAGCCCAAGT 34218  
Qy 804 ACAGCATGAGGCTCGGGGACAAAGAGGCTGCTGAGCGACGATGGAATTTTCACTTGGCT 863  
Db 34219 CACCAGGGGCTCCAGAGATGTGAGAGCCGCTGACCCAAATGCTGACGTTCTCCCTGGGCT 34278  
Qy 864 GTTTTGCATTCCTCTCTTGAAGAAGACTATCCAGAGCATGAGAGCAGCTGG 923  
Db 34279 GGTTCGCTCAACCCATTTTGAAGAAGGAGACTATCTGACACCATGAAGTGAAGTGG 34338  
Qy 924 GCGAGAGG 931  
Db 34339 GGAACAGG 34346

RESULT 11  
US-10-767-701-8210  
; Sequence 8210, Application US/10767701  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(5353)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 8210  
; LENGTH: 689  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS33734\_1  
US-10-767-701-8210

Query Match 6.3%; Score 125.2; DB 6; Length 689;  
Best Local Similarity 56.4%; Pred. No. 1.6e-22;  
Matches 299; Conservative 0; Mismatches 218; Indels 12; Gaps 3;  
Qy 73 GCTAGCTCTGCCCAACGACTTTGATGGGCTTGCACAGCGCCCTACAGATCGAAG 132  
Db 173 GCAGGATTTTCCAGGGCTTCGTTGGGGCCGGACATCAGCTTATCAGTGAAG 232  
Qy 133 CGCCGTAAGAAGAGTGGCGGCGCCCTCATCTGGGACACCTACTGCACTGGAGCC 192  
Db 233 GGGCAACGATGAGATGAGAGAGCCCAAGCATATGAGACACGTTTACTATGACGGGAG 292  
Qy 193 ATGGCGCACCAAGCGGCGCAACGGGATGTGGCTTGGATCATACACCGCTACGATGA 252  
Db 293 GATGCCGCAAAAGCACT--GGCATCTAGGGGACGACGCGCTTACCACAAAATACAGGA 349  
Qy 253 GGAATTGATCTTGAACGAGTACGGCGAAGGCTTACCGCTTCTCTTGTGTGTC 312  
Db 350 AGATGTGAATGTGATGATGATGCTGAGCGCTGAGCGCTTCCATTTCTTGTGTC 409  
Qy 313 GCGGATATTCCCTCGGCGGACGCTGATCCGCTCAACGAGAGGAAATTGATTTA 372  
Db 410 CAGGCTTCTTCAAGAGGAAG-----AGGACCATCAACCCCAAGGGCTTTCAGTATTA 463  
Qy 373 CAGCAACGATTTGAGCGCCCTTGAAGCGGGGATATACGCGCTTGGGCTTTGTACA 432  
Db 464 CAACCACTTATCATATGATGATCAATCAACGGGGAATGAGATACAGTGAACCTGTACCA 523  
Qy 433 CTGGGATCTGCTCAGCGCTTCAAGTCTATGAGAGGCTGCTCAACGTGAAGAGAT 492  
Db 524 CTGGGATTTCTCTAGATCTCTGAGAGAGATACATCAAGGCTGCTGAGCCCGAGGTGT 583  
Qy 493 CCAAGCTGAGCTTGAAGCGGTATGCGAGGTTGTCTTTGAACGTTTGGGAGCGAGTCA 552  
Db 584 ---GAGAGCTTCAACGCGCTGACGCGGACGTGTCTCCGCGGAGTTTGGCGCAGCGGTGAA 640

Qy 553 GAACGTGATCATCATCAACGAGCCCTGATTTAGGCCCATCTATGATGA 600  
Db 641 GCACTGGACGACCATGAGACGAGCCGAAACGTGATCTCATCGCGGCTTA 688

RESULT 12  
US-10-767-795-4264  
; Sequence 4264, Application US/10767795  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(5353)B  
; CURRENT APPLICATION NUMBER: US/10/767,795  
; CURRENT FILING DATE: 2004-01-30  
; NUMBER OF SEQ ID NOS: 117596  
; SEQ ID NO 4264  
; LENGTH: 649  
; TYPE: DNA  
; ORGANISM: Goessypium hirsutum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C301\_1  
US-10-767-795-4264

Query Match 6.0%; Score 119.4; DB 6; Length 649;  
Best Local Similarity 54.5%; Pred. No. 4.8e-21;  
Matches 285; Conservative 1; Mismatches 227; Indels 10; Gaps 2;

Qy 79 TCTGCCCAACGACTTTGATGGGCTTGCACAGCGCCCTACAGATCGAAGCGGCT 138  
Db 137 TTTTCGAGAGGATTTGTGTGGAACCTGCACTTACGCTTATCAAGTGAAGGATGGC 196  
Qy 139 CAAAGAGGTGGCGCGCCCTCATCTGGGACAGCTACTGCCACCTGAGCCATGCG 198  
Db 197 TAAATAAGCGCGCGGAGGACCTTGCATTTGGGATGTTTATGTTAAACACAGGCCATAT 256  
Qy 199 CACCAACGGGCCAAGCGGATGTGGCTTGCATCTACACCGCTAGAGATGAGCTT 258  
Db 257 TGCTAATATGATGATGCTGACGCTGTGACACGATACACCATTAACAAGAGATGT 316  
Qy 259 TGATCTCTTGAACGAGTACGCGCAAGGCTTACCGCTTCTCTTGTGCTGCGAGAT 318  
Db 317 AGATTTGTTGGCAATTTTCAATTTGATTTCTTATCGTTCTCAATTTGATGATGAT 376  
Qy 319 CATTTCCCTTGGCGGACGCTGATCCGCTCAACGAGAGGAAATTGATTTTACAGCA 378  
Db 377 CTTCAGAGGGTG-----TTGGAAGATTAATTGGAGGAGGATGATTATTAACAG 430  
Qy 379 ACTGATGAGCGCCCTGTGAGCGGGGATACAGCCCTTGGGATCTTTGATACACTGGGA 438  
Db 431 GTTGATCAATTAATCTGTTAAGAAAGATTAATCTGATGAGAACTTGTACATTAATGA 490  
Qy 439 TCTGCTCAGGCGCTTCAACGATGCTATGAGAGGCTGCTCAAGTGAAGAGTCCAGCT 498  
Db 491 TCTCCCTCAAGCTTCTTCAAGAGATCAATGAGGTTCTTGAACCGTCAA-----TTGTGA 546  
Qy 499 GGAATTGAGCGGTATGCGAGGTTGTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTG 558  
Db 547 AAATATGCTGATATATGCAAAAGTTTGTTCAAAACATTTGGGGATTTAAATGCAAAACTG 606  
Qy 559 GATCAGCATCAACGAGCCCTGATTCAGGCACTATGATAT 601  
Db 607 GTTTAATTAATGAACCCAGAAATCGGGGTGTTTGTGTTT 649

RESULT 13  
US-10-767-701-12773  
; Sequence 12773, Application US/10767701  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 12773  
LENGTH: 716  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS25405\_1  
US-10-767-701-12773

Query Match 5.8%; Score 115.4; DB 6; Length 716;  
Best Local Similarity 56.6%; Pred. No. 5.3e-20;  
Matches 237; Conservative 0; Mismatches 176; Indels 6; Gaps 1;

QY 81 TGCCCAAGACTTGAATGAGGCTTCCGACGCGCGCTTACAGATCGAAGCGCGCTCA 140  
DB 281 TCCCAAGGGTGTGTGTCGGAAGCGGACGTGCGCTTACAGATCGAAGCGCGCGCT 340  
QY 141 AAGAAGTGCGCGCGCGCTTCCATCTGGACACGTACTGCCACTGAGCCATCGCGCA 200  
DB 341 CCACCAACGCGCGGACCTTCATCTGGATTCATTCGCGCACGTCCAGAAATATTG 400  
QY 201 CCAACGCGCGCAAGCGGCGATGCGCTTCCGATCTACACGCTTACAGATCGAAGCTT 260  
DB 401 CGGGAATCAAAATGGAACCTTGATGATCAATCAATCAATCAATCAATCAATCAAT 460  
QY 261 ATCTTGAACCAAGTACGCGCGCAAGCGCTTCCCTTGTGCTGTGCGGATCA 320  
DB 461 ATCTGATGAAGATTTGATTTGATGCTACCGCTTTTCAATCGTGTCGAGATCT 520  
QY 321 TTCCCTTGGCGGAGCTGATCCGTCACGAGAGGGAATGATTTTACAGCAAC 380  
DB 521 TCCGAGATGCGG-----AGGGAAGATCAATCCAGAGGTGATGATTAATAATC 574  
QY 381 TGATTTGACCGCTTGTGAGGCGGCTATCAGCGCTGCGTGTGATTTGACGATCGATC 440  
DB 575 TGATTAATATCTGCTTCAAGAAAGCATGACTCTTACATCAACCTTTACCATATGATC 634  
QY 441 TGCTCAGCGCTTACGATCGCTATGAGGCTGCTCAAGTGGAGAGGCTCCAGCTG 499  
DB 635 TTCTCTTGGCTTGAAGAAATATGAGGCTTAAAGCGCTTAAGATGCGGACCTG 693

RESULT 14  
US-10-767-701-26695  
Sequence 26695, Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 26695  
LENGTH: 637  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: 30977777  
US-10-767-701-26695

Query Match 5.8%; Score 114.8; DB 6; Length 637;  
Best Local Similarity 58.4%; Pred. No. 7.2e-20;  
Matches 222; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 77 GCTCTGCCAAGCACTTGAATGAGGCTTCCGACGCGCGCTTACAGATCGAAGCGCGC 136  
DB 256 GCGTTCCCAAGGGGTTTCATCTTCGGAAGCGGACGTCGCGCTTCCAGTCTGAGGCGCG 315  
QY 137 GTCAAGAAGTGGCGCGCGCGCTTATCTGGACACGTAATGCGACCTGAGGCTATCG 196  
DB 316 GCACGTCGCGCGCGCGCGCGCGCGCATCTGGATCTCTTCTGTCACACCGCGGAAG 375  
QY 197 CGACCAAGCGCGCGCGCGCGCGCGCTTGGCTTGCATCTACACCGCTAGATGAGAC 256  
DB 376 ATGCTGAAGACGGAACGAGATGTTACAGATGAATATCATGCTCAAGAGAT 435  
QY 257 TTGATCTTTGACCAAGTACGCGCGCAAGCGCTTCCCTTCTGTGTGCTCGCG 316  
DB 436 GTTATCTCATGAAGAACCTTAATTTGATGACATACCGGTTTCAATCTCTGCTCAGG 495  
QY 317 ATCATTCCTCGCGCGCGCGCGCGCGTATCCGCTCAAGAGAGGAATTGATTACAGC 376  
DB 496 ATCTTCCAGATGCGGAAGG-----AAAGTATGAAGAGGATGACGATTTACAC 549  
QY 377 AAATGATGACGCGCTTGTGAGGCGGCTATCACGCTTGGTGAATTTGATCACTG 436  
DB 550 AATCTTATGATCAATGATTTAAGCAAGTCTTACTCTTAAGCAACCTTAACCATAT 609  
QY 437 GATCTGCTCAGCGCTTCA 456  
DB 610 GATCTTCACTTGGCTTCA 629

RESULT 15  
US-10-767-795-4474  
Sequence 4474, Application US/10767795  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,795  
CURRENT FILING DATE: 2004-01-30  
NUMBER OF SEQ ID NOS: 117596  
SEQ ID NO 4474  
LENGTH: 558  
TYPE: DNA  
ORGANISM: Gossypium hirsutum  
FEATURE:  
OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C30142\_1  
US-10-767-795-4474

Query Match 5.5%; Score 109; DB 6; Length 558;  
Best Local Similarity 58.7%; Pred. No. 2.1e-18;  
Matches 225; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 77 GCTCTGCCAAGCACTTGAATGAGGCTTCCGACGCGCGCTTACAGATCGAAGCGCGC 136  
DB 107 GATTTCCCTCCCACTTGTCTTGTGCTTCTTCACTTCTTCAATGATGAGGAGCT 166  
QY 137 GTCAAGAAGTGGCGCGCGCGCGCGCTTATCTGGACACGTAATGCGACCTGAGGCTATCG 196  
DB 167 GTGAACGAAGTGGCAAGGTAAGAAATATGAGATCTTCTTCTCACATTTGAAGAAA 226  
QY 197 CGACCAAGCGCGCGCGCGCGCGCGCTTGGCTTGCATCTACACCGCTAGATGAGAC 256  
DB 227 ATGCTGAAGACGAATGCTGATGTTGCACTGATCTTACACAGTACAGAGAT 286  
QY 257 TTGATCTTTGACCAAGTACGCGCGCAAGCGCTTCCCTTCTGTGTGCTCGCG 316  
DB 287 ATAGGCTTATATCGAGTTAGGCTTCAAGCTTACAGATTTTCCATATCATGCGCTCGT 346  
QY 317 ATCATTCCTCGCGCGCGCGCGCGCTGATCCGCTCAAGAGAGGAATTGATTACAGC 376  
DB 347 ATTTTCCAGATG--GTTTGGAAACCAAGTATATGAGAGGAATTTGATTTTACAC 404

Qy	377	AAACTGATTGACGCGCCCTGTTGAGCGGGGTATCACGCGCTTGGGTGA	CTTTGTACCACTGG	436
Db	405	AATGTGATCGATGCGCTT-TTTGAAAGGTAATTGAGCGCCCTTGTGACATTATACCA	TTGG	463
Qy	437	GATCTGCGCTTCAGGCGCTTCACGA		459
Db	464	GATTCTTCCTTGCATCTCCATGA		486

Search completed: March 29, 2004, 02:56:06  
Job time : 176.02 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2004, 20:57:07 ; Search time 3903.5 Seconds  
(without alignments)  
1130.850 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455  
Sequence: 1 atgccgcagctcgtcgtctc.....gggttaagtggcgcatcaa 1455

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST.\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_pro:\*  
26: em\_gss\_rtd:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	702.2	48.3	773	14	CB904679 trico39xe
2	701	48.2	703	14	CF875727 trico39xe
3	461.2	31.7	584	14	CF884351 trico39xe
4	387	26.6	1994	9	AA415086 Mg0020 RC

5	201	13.8	2031	11	AY109400
6	195.2	13.4	786	14	CB677158
7	193	13.3	788	14	CB673495
8	190.4	13.1	806	14	CB642881
9	179.2	12.3	889	14	CD438244
10	175.6	12.1	480	14	CB688082
11	175	12.0	1182	11	AY106991
12	171.8	11.8	662	14	CF880605
13	171.8	11.8	752	14	CB907493
14	169.8	11.7	650	10	BE777259
15	162.6	11.2	613	9	AJ436490
16	158.6	10.9	860	14	CF666831
17	157.8	10.8	610	13	BQ110253
18	157	10.8	825	14	CK200642
19	156.2	10.7	1136	14	CK209455
20	155.8	10.7	708	13	CA145286
21	155.6	10.7	673	14	CD482402
22	155.6	10.7	674	14	CD458246
23	154	10.6	786	14	CB634193
24	153.4	10.5	792	14	CB633785
25	153.4	10.5	808	14	CB633785
26	153.2	10.5	729	13	CA139881
27	153	10.5	1126	14	CK161716
28	152.4	10.5	788	14	CF554428
29	151.8	10.4	728	14	CF473575
30	151.6	10.4	723	12	BM605139
31	151.6	10.4	766	14	CB643857
32	151.6	10.4	821	14	CB634605
33	151.2	10.4	564	14	CB641046
34	151.2	10.4	636	9	AV913799
35	151	10.4	676	12	BM621029
36	151	10.4	805	14	CB677500
37	150.8	10.4	646	12	BJ235347
38	150.8	10.4	829	14	CK201813
39	150.6	10.4	829	14	CB663657
40	150	10.3	592	14	CA193361
41	150	10.3	760	14	CB643189
42	150	10.3	794	14	CB654045
43	150	10.3	806	14	CB674932
44	150	10.3	826	14	CB648777
45	150	10.3	856	14	CB668311

## ALIGNMENTS

RESULT 1  
LOCUS CB904679  
DEFINITION trico39xe13 T. reesei mycelial culture, Version 3 apr11 Hypocrea  
ACCESSION CB904679  
VERSION CB904679.1 GI:30119337  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
REFERENCE  
AUTHORS  
Foreman, P.K., Brown, D.E., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., Englund, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,  
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.  
Transcriptional regulation of biomass-degrading enzymes in the  
filamentous fungus Trichoderma reesei  
J. Biol. Chem. 278 (34), 31988-31997 (2003)  
MEDLINE 22803314  
PUBMED 12788920  
TITLE  
COMMENT  
Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817

Email: Pforeman@genecor.com  
Seq primer: LT-F1 primer.  
Location/Qualifiers  
source

1..773  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric039x13"  
/dev\_stage="mycelia"  
/clone\_lib="T. reesei mycelial culture, Version 3 April"  
/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## ORIGIN

Query Match 48.3%; Score 702.2; DB 14; Length 773;  
Best Local Similarity 92.8%; Pred. No. 5.1e-129;  
Matches 707; Conservative 1; Mismatches 54; Indels 0; Gaps 0;

279 GGATCCCGTCAACGAGGAGGAATTGATTTCACGAACTGATGACGCCCTGTGAG 338  
12 GNNNCTGCTGCTGTAATNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNG 71  
339 GCGGGGTATCAAGCCTTGGTGAATTTGACCTGGGATGCTGCTCAGCGCTTCAAGA 398  
72 GCGGGGTATCAAGCCTTGGTGAATTTGACCTGGGATGCTGCTCAGCGCTTCAAGA 131  
399 TCGCTATGAGAGGCTGCTCAAGTGAAGAGAGTCCAGTGAATTTGAGCGGTATGCGAG 458  
132 TCGCTATGAGAGGCTGCTCAAGTGAAGAGAGTCCAGTGAATTTGAGCGGTATGCGAG 191  
459 GTTGTGCTTTGAAGCTTTGGGAGCCGAGTCCAGAACTGATCAACATCAAGMCCCTG 518  
192 GTTGTGCTTTGAAGCTTTGGGAGCCGAGTCCAGAACTGATCAACATCAAGMCCCTG 251  
519 GATTGAGCCATCTATGATATGCCAAGCGGACGAAGCGCCCGGAGAGAGACATTTAA 578  
252 GATTGAGCCATCTATGATATGCCAAGCGGACGAAGCGCCCGGAGAGAGACATTTAA 311  
579 CAAGCATCTCCACGAGGCAACACTGACGAGCCGAGTCCGCTGGAAGGCGCAGAT 638  
312 CAAGCATCTCCACGAGGCAACACTGACGAGCCGAGTCCGCTGGAAGGCGCAGAT 371  
639 CATGAGCCATGCGCGCGCGCTGCTTACAGCAGGAGCTTTGCGCCCTGCAAAAAGG 698  
372 CATGAGCCATGCGCGCGCGCTGCTTACAGCAGGAGCTTTGCGCCCTGCAAAAAGG 431  
699 CCAAGTGGGATCTGCTCAAGCGGACTATGAGCCCTGGGACAGCAATGAGCCTCG 758  
432 CCAAGTGGGATCTGCTCAAGCGGACTATGAGCCCTGGGACAGCAATGAGCCTCG 491  
759 GGAAGAAGAGGCTGAGAGGAGATGGAATTTCAATGGCTGGTTGCAATCCCAT 818  
492 GGAAGAAGAGGCTGAGAGGAGATGGAATTTCAATGGCTGGTTGCAATCCCAT 551  
819 CTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGAGCTGGGAGAGGCTTCCAGC 878  
552 CTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGAGCTGGGAGAGGCTTCCAGC 611  
879 CTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGAGCTGGGAGAGGCTTCCAGC 938  
612 CTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGAGCTGGGAGAGGCTTCCAGC 671  
939 TTACTACATCCAGAGTTCGCGGCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 998  
672 TTACTACATCCAGAGTTCGCGGCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 731  
999 CCGGCGCATCCATGAGACCAAGAGAAATAAGAGCGGAGAGCC 1040  
732 CCGGCGCATCCATGAGACCAAGAGAAATAAGAGCGGAGAGCC 773

RESULT 2  
CF875727  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CF875727 703 bp mRNA linear EST 31-OCT-2003  
tric039x13.b1.T.reesei mycelial culture, Version 6 October 2003  
Hypocrea jecorina cDNA clone tric039x13, mRNA sequence.  
CF875727  
GI:38130409  
EST.

Hypocrea jecorina (anamorph: Trichoderma reesei)  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

REFERENCE  
1 (bases 1 to 703)  
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D., and Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and Dean, R.A.

Analysis of the protein processing and secretion pathways in a Trichoderma reesei EST dataset  
Unpublished (2003)  
Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu

Seq primer: LT-F1 primer.  
Location/Qualifiers

1..703  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="tric039x13"  
/dev\_stage="mycelia"  
/clone\_lib="T. reesei mycelial culture, Version 6 October 2003"  
/note="Vector: pREP3; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## FEATURES

source

## ORIGIN

Query Match 48.2%; Score 701; DB 14; Length 703;  
Best Local Similarity 99.7%; Pred. No. 8.5e-129;  
Matches 701; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

338 GGGGGGTATCACGCCCTTGGGTGACTTTGTACCACTGGGATCTGCTCAGGCGCTTCAAG 397  
1 GGGGGGTATCACGCCCTTGGGTGACTTTGTACCACTGGGATCTGCTCAGGCGCTTCAAG 60  
398 ATGCTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCTGACTTGAAGCGGTATGCGA 457  
61 ATGCTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCTGACTTGAAGCGGTATGCGA 120  
458 GGTGTGCTTTGAAGCTTTGGGAGCCGAGTCCAGAACTGATCACCATTCAAGMCCCT 517  
121 GGTGTGCTTTGAAGCTTTGGGAGCCGAGTCCAGAACTGATCACCATTCAAGMCCCT 180  
518 GGAATGAGGCACTTATGATATGCCAAGCGGAGCAAGCGCCCGGAGAGAGAGAGATTA 577  
181 GGAATGAGGCACTTATGATATGCCAAGCGGAGCAAGCGCCCGGAGAGAGAGATTA 240  
578 ACAAGACTCCACGAGGCAACACTGACCTGAGCCGTGCTGCTGAAAAGGCCCAAGA 637  
241 ACAAGACTCCACGAGGCAACACTGACCTGAGCCGTGCTGCTGAAAAGGCCCAAGA 300  
638 TCATGAGCATGCGCGCGCGCTGAGCCGTCTACAGCAGGAACTTTGCGCCCTGCAAAAAG 697  
301 TCATGAGCATGCGCGCGCGCTGAGCCGTCTACAGCAGGAACTTTGCGCCCTGCAAAAAG 360  
698 GCGAGATGGGATCTGCTCAAGCGGAGACTATGATGAGCCCTGGGACAGCAATGAGCCTC 757  
361 GCGAGATGGGATCTGCTCAAGCGGAGACTATGAGCCCTGGGACAGCAATGAGCCTC 420

QY 758 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGGCTGATTTGCCAATCCCA 817  
| | | | |  
Db 421 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGGCTGATTTGCCAATCCCA 480  
| | | | |  
QY 818 TCTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGACGCTGGCGAGAGGCTTCCAG 877  
| | | | |  
Db 481 TCTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGACGCTGGCGAGAGGCTTCCAG 540  
| | | | |  
QY 878 CCTCTACTCCCGGAGACTTTGGCATCTCTCAATGCCGAGAGACCGACTTTACGGCATGA 937  
| | | | |  
Db 541 CCTCTACTCCCGGAGACTTTGGCATCTCTCAATGCCGAGAGACCGACTTTACGGCATGA 600  
| | | | |  
QY 938 ATTACACACATCCGAGTTCGCGCGACCTTAGACGGTCCGTCGCCGAGACGAGATATC 997  
| | | | |  
Db 601 ATTACACACATCCGAGTTCGCGCGACCTTAGACGGTCCGTCGCCGAGACGAGATATC 660  
| | | | |  
QY 998 TCGGCGCCATCCATGAGCACACAGAGAATGAAGACGGCAGCCC 1040  
| | | | |  
Db 661 TCGGCGCCATCCATGAGCACACAGAGAATGAAGACGGCAGCCC 703  
| | | | |

RESULT 3  
CF884351 584 bp mRNA linear EST 31-OCT-2003  
LOCUS trico39xel3.b12 T. reesei mycelial culture, Version 6 October 2003  
DEFINITION Hypocrea jecorina cDNA clone trico39xel3, mRNA sequence.  
ACCESSION CF884351  
VERSION CF884351.1 GI:38139033  
KEYWORDS EST.  
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)  
ORGANISM Hypocrea jecorina  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.  
1 (bases 1 to 584)  
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,  
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and  
Dean, R.A.  
Analysis of the protein processing and secretion pathways in a  
Trichoderma reesei EST dataset  
JOURNAL Unpublished (2003)  
COMMENT Contact: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu  
Seg primer: Tr-F1 primer.  
Location/Qualifiers  
FEATURES  
source  
1..584  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/clone="trico39xel3"  
/dev\_stage="mycelial"  
/clone\_lib="T.reesei mycelial culture, Version 6 October  
2003"  
/note="Vector: pREP3Y, Site 1: Not I/Sal I, Mycelial  
culture grown from 24 hrs to 6 days with varying Carbon  
and Nitrogen sources and concentrations."

ORIGIN  
Query Match 31.7%; Score 461.2; DB 14; Length 584;  
Best Local Similarity 94.0%; Pred. No. 3,3e-81;  
Matches 489; Conservative 1; Mismatches 29; Indels 1; Gaps 1;  
QY 338 GGGGGGATACAGCGCTTGGGTGACTTTGATACCACTGGGATTCGCTCAGACGGCTTACG 397  
| | | | |  
Db 1 GGGGGGATACAGCGCTTGGGTGACTTTGATACCACTGGGATTCGCTCAGACGGCTTACG 60  
| | | | |  
QY 398 ATCGTATGAGAGCTGCTCAACGTGGAAGAGTCCAGTGCATTTGACGGGTATGCGA 457  
| | | | |

Db 61 ATCGTATGAGAGCTGCTCAACGTGGAAGAGTCCAGTGCATTTGACGGGTATGCGA 120  
| | | | |  
QY 458 GGTGTGCTTTGAACGTTTGGGAGACCGATCCAGAACCTGGAATCACATCAAGACGCC 517  
| | | | |  
Db 121 GGTGTGCTTTGAACGTTTGGGAGACCGATCCAGAACCTGGAATCACATCAAGACGCC 180  
| | | | |  
QY 518 GGAATGAGCCATCTATGATATGCGACCGGACGAAACCCCGGCGAGAGAGCATTA 577  
| | | | |  
Db 181 GCATTAGGCCATCTATGATATGCGACCGGACGAAACCCCGGCGAGAGAGCATTA 240  
| | | | |  
QY 578 ACAAGCACTTCACCGAGGCGAACACTGCACTGAGCGTGTGCTGCTGGAAGGCCAGA 637  
| | | | |  
Db 241 ACAAGCACTTCACCGAGGCGAACACTGCACTGAGCGTGTGCTGCTGGAAGGCCAGA 300  
| | | | |  
QY 638 TCATGAGCCATGCGCGCGCGCGTGGCCGCTACAGAGGAGACTTCCGCCCTGCAGAAAG 697  
| | | | |  
Db 301 TCATGAGCCATGCGCGCGCGCGTGGCCGCTACAGAGGAGACTTCCGCCCTGCAGACG 360  
| | | | |  
QY 698 GCCAGATCGGCATCTGCTCAACGCGGACTATGAGCCCTGGGACAGCAATGAGCCTC 757  
| | | | |  
Db 361 GCCAGATCGGCATCTGCTCAACGCGGACTATGAGCCCTGGGACAGCAATGAGCCTC 420  
| | | | |  
QY 758 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGGCTGATTTGCCAATCC 816  
| | | | |  
Db 421 GGGACAAGAGGCTGCTGAGCGACGAGATGGAATTTTCACATTGGCTGATTTGCCAATCC 480  
| | | | |  
QY 817 ATCTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGC 856  
| | | | |  
Db 481 ATCTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGC 520  
| | | | |

RESULT 4  
AA415086 1994 bp mRNA linear EST 09-DEC-1999  
LOCUS MG0020 RCW Lambda Zap Express Library Magnaporthe grisea cDNA clone  
DEFINITION RCW20 similar to Beta-Glucosidase (EC 3.2.1.21), mRNA sequence.  
ACCESSION AA415086  
VERSION AA415086.1 GI:12537251  
KEYWORDS EST.  
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetidae; Incertae sedis; Magnaportheaceae; Magnaporthe.  
1 (bases 1 to 1994)  
Wu, S.-C., Bernstein, B.D., Davull, A.G. and Albersheim, P.  
Expressed sequence tags of the rice blast fungus grown on rice cell  
walls  
Unpublished (1997)  
JOURNAL Contact: Sheng-Cheng Wu  
COMMENT CCRC  
University of Georgia  
220 Riverbend Road, Athens, GA 30602-4712, USA  
Tel: 706 542 4446  
Fax: 706 542 4412  
Email: wus@ccrc.uga.edu  
Fully sequenced  
Insert Length: 1994 Std Error: 0.00.  
Location/Qualifiers  
FEATURES  
source  
1..1994  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/strain="CP987"  
/db\_xref="taxon:148305"  
/clone="RCW20"  
/sex="Mati-1 hermaphrodite"  
/tissue\_type="mycelium"  
/dev\_stage="Day 5 post-inoculation"  
/clone\_lib="RCW Lambda Zap Express Library"  
/note="Vector: pBluescript excised from Lambda Zap  
Express; Site 1: EcoRI; Site 2: XhoI; Day 5  
post-inoculation mRNAs prepared from Magnaporthe grisea  
grown at 23C in the dark with constant gyrotory shaking

100 rpm in Vogel's minimal medium containing 0.5% isolated rice cell walls as the sole carbon source. Library provided by Sheng-Cheng Wu. Sequences were processed by one of two methods. Where a full-length alignment to the M. grisea genome sequence was available, the EST sequence was trimmed according to the alignment, otherwise sequence quality was assessed using phred/Phrap version 991019 and trimmed according to phd files (0.05) and for vector segs."

# ORIGIN

Query Match 26.6%; Score 387; DB 9; Length 1994;  
Best Local Similarity 57.0%; Pred. No. 3.1e-66;  
Matches 806; Conservative 1; Mismatches 591; Indels 15; Gaps 5;

17 CTGCGCCCAAGACCTTGAATGAGGCTTCCGACAGCGCGCTTACCAATGCAAGCGCG 76  
290 CTCTCCCAAGATTTCTTTGGGCTTCCGACAGCGCTCAATACAGATCCAGAGGTGCTA 349  
77 TCAAGAAGGTGGCCGCGCCGCTCCATCTGGACACGTAATGCGACCTGGAGCCATGCG 136  
350 TCGACAAGAGTGTGTGGGCGCTCCATTTGGGATCTTACTGCAATTCGGGTAAAG 409  
137 GCACCAAGCGGCCAAGCGCATGTGCTTGCATCACTACCAAGCGCTACGATGAGACT 196  
410 TCGCAGACGGCAGCTCGGCTGTGACGGCATGCGATCACTACACAGACGAGGAGCA 469  
197 TTGATCTTGGACCAAGTAAGCGGCAAGGCTTCCGCTTCTTGTGCTGGCGGA 256  
470 TTGACCTGCTCAATGCTGTGTGTGTCACCAAGTATCCCTTTTCATCTGCTGGTCGAGA 529  
257 TCATTTCCCTCGCGGAGAGCTGATCCGCTCAACGAGAGGGAATTTGATTTACAGA 316  
530 TTATCCCATGTGGTGGCGTAATGACCCCATCAACGAAAGGAGTGAACACACTACGTCA 589  
317 AACTGATTAAGCGCTTGTGAGGGGGTATCAAGCTTGGGTGACTTTGATCACTGGG 376  
590 AGTTGTAAGACGACTCTGAGGGCCGCAATCAAGCCATCATCACCTCTTCCACTGGG 649  
377 ATCTGCTCAGGCGCTTACGATGCTATGAGGCTGAGCTCAACGTGAAAGAGGTCCAGC 436  
650 ATCTGCGGATGATTTGACAGAGCGTACGCTGCTGTAACAGGAAAGATTCCTT 709  
437 TGAATTTGAGCGGATGATGAGGTTGCTTGAACGTTTGGGAGCCGATCCAGAACT 496  
710 TGAATTTGAGCACTATGACGCGTCACTGTTCAAGGCGATTTCCCAAGTG---CAAACCT 766  
497 GGAATCAACATCAACGACCTTGAATTTAGGCACTTATGATATGCAACGCGGACAGC 556  
767 GGAATCACTTCAACGAGCGCTGTGCTGCTCAATCTCGCTTACAGGTTGCTCAGTTTG 826  
557 CCGCGGACGAGGAGCATTAACAGACATCCACGAGGCGCAACATGCACTGAGCGCT 616  
827 CGCGTGGCGCTGTCCGACCGGACGAAAGTCCCTGTCCGCGACAGCTTCGGGAGCCCT 886  
617 GCGTCTGTAAGAGCCCAATATAGGCAATGCGCGCGCTGCGCTGTACAGCAGAGG 676  
887 GGAATTTGGGCGCAACCTTCTGTGTGGCCACCGTGTGCGCTCAAGGTGTACAGGAAAG 946  
677 ACTTTGGCCCTCGCAAAAGGCGCAATGGGCATCTCGCTCAACGCGGATCTATGAGC 736  
947 AATTCAAGGCGGACAGAGGAGTGTGATGATCAAGTTGAATGGGAGCGCACTTCC 1006  
737 CTTGGGACAGCAATGAGCTTGGGACAGAGGCTGTGAGCAGCGATGGAATTTCA 796  
1007 CTTGGGATCCCGAAGACCCAGGAGGTGATGACAGCAACCGCAATGATGATTTGCCCA 1066  
797 TTGGCTGTTTGGCAATCCATCTTTTGAAGAAGACTATCCAGAGCATGAAGAAGC 856  
1067 TCTGCTGTTTGGCGACCCCATCTACTTTGGCAG---TATCCAGTCTCGATCGTAAGC 1123  
857 AGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCAATCTCAATGCGGAG 916

Db 1124 AGCTTGGCGACCGCTTCCCACTTTTACGAGAGGAGAAAGCTTTAAGTCAA---GGCTT 1180  
Qy 917 AGACCGACTTCTACCGCATGAATTAATACACATCCAGTTTGGCGGCACCTTACAGCTC 976  
Db 1181 CAACGACTTTTATGGAATGAATGCTACACGCGCAACTCATCAAGGACCAAGAAAGAG 1240  
Qy 977 CCGTCCCGGAGAGGACTATCTGCGGCGCTTCCATGACACCGAGAAATTAAGACGGCA 1036  
Db 1241 AGCCGCGCGAGAGCAATTAATCTTGGAAACCTGGAAACAGTTTCTCAACACAGCGCGC 1300  
Qy 1037 GCCCGCTTGGCGAGAGAGAGCGCGCTCGCTGCTGCTCTCCGCGCATGCTTCCGGA 1096  
Db 1301 AATGCAATCGCGCGGAGAGACACATCAACCTGCTTGTGCTCAACGACAGAGTTCCGAG 1360  
Qy 1097 AGCATCTTCCCGCGGTGTACGCGCTTGTACGCGCAAGCTC---ATCTACATCAAGAGACG 1153  
Db 1361 AGCTGCTGCTTGTGCTCAGCAAGCTTACCAATCAACCCCAAGATCTTGTGTACGAGAAATG 1420  
Qy 1154 GATGCCCGTGCCTGGAGAGAGACATGACGTCGAGAGAGCGCGCTCAACGACCCCTTCC 1213  
Db 1421 GCACGTGCTCAAGGCGGAGACGATATGCGCTCGAAGAAATCTTCCAGATGACTTCC 1480  
Qy 1214 GCATCGGTAATTTGACTGCACTTGGACTGATTTTCAAGGCGCATTAACCGAGAGCGG 1273  
Db 1481 GGGTCCAGTACTACGACGATTAATGTCAGAGCTTTGGCCAGGCTTATTCGAGGATGTG 1540  
Qy 1274 TCGTCTCAAGGAGTACTTTGCTGGGCGTTGTGCTATTAATCTTGAATGCTCAGATGCT 1333  
Db 1541 TAAACGTCCGCGGATCTCAGCTGTGTGATGACAACTTTGATGGCGCGAGGGCT 1600  
Qy 1334 ACAGACCAAGATTCGCGCTCAGCTTCAAGACTAC---ACCAACCTCAAGCGGAGCGCA 1390  
Db 1601 ACAGACCAAGGTTTGTGCTCACTTTGTGACTACGAGAAACGCGCAAGGCTTACCCA 1660  
Qy 1391 AGAAGTTCGCTGCTCTCAAGGACATGTTTG 1423  
Db 1661 AGAAGAGCGCCAAAGCAATGAAGCGTGTGTTG 1693

RESULT 5  
AY109400 2031 bp mRNA linear HTC 17-OCT-2002  
LOCUS  
DEFINITION Zea mays CL1052.1 mRNA sequence.  
ACCESSION AY109400  
VERSION AY109400.1 GI:21213111  
KEYWORDS  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 2031)  
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes  
Unpublished (2002)  
2 (bases 1 to 2031)  
JOURNAL  
REFERENCE  
AUTHORS  
TITLES  
COMMENT  
Direct Submission  
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
If you are interested in getting corresponding physical clones, these are publicly available in getting from ZmDB and may be found by BLAST searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.  
location/Qualifiers  
1..2031  
/organism="Zea mays"  
/mol\_type="mRNA"

/db\_xref="MaizeDB:629921"  
/db\_xref="taxon:4577"  
/clone\_id="Maize Mapping Project/Dupont Cornsensus  
Library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed Dupont contigs; this resource was  
assembled by Dupont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

## ORIGIN

Query Match 13.8%; Score 201; DB 11; Length 2031;  
Best Local Similarity 51.3%; Pred. No. 2,7e-29;  
Matches 705; Conservative 1; Mismatches 632; Indels 36; Gaps 9;

55 GCCTACCATGAGAGCGCGCTCAAGAGAGTGGCGCGCGCGCTCCATCTGGGACG 114  
Db 5 GCGTACCAAGTTACGAGGGCGCGCTCCACCAAGCGCGCGCGCGCTCCATCTGGGATTCA 64  
Qy 115 TACGCGCAGTGGAGCATGGCGACCAAGCGCGCGCGCGCGATGGCGCTTGGGATC 174  
Db 65 TTGCGGACGCTCCAGGAAATATTTGACGGAAATCAAAATGAGAGCTTGCAGTGAATCA 124  
Qy 175 TACCAAGCTACGATGAGACTTGTATCTTGAACCAAGTACGGCGCAAGGCGCTACCGC 234  
Db 125 TACCATGGCTACAAAGAAAGTGCATCTCATGAAGTTGAACTTTGATGCTTACCGG 184  
Qy 235 TTCTCTTGTGTGTGGCGGATATTCCTCGGCGGCGAGCTGGATCCGCTCAACGAG 294  
Db 185 TTCTCAATCTCATGTCCAGGATCTCCGCGATGGCGAAGG-----AAAGTCAATCCA 238  
Qy 295 GAGGAAATGAGTTTTCAGAAACTGATMGACCCCGCTTGAAGCGGCGGATACCGCT 354  
Db 239 GAAAGTGAAGGATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 298  
Qy 355 TGAGTGAATTTGATCACTGGGATCTGGCTCAGCGCTTCCAGATCCGCTATGAGAGCTG 414  
Db 299 TACATCAACCTTTACCACTATGATCTTCTTGGCTGGAAGAAATATGAGAGGCTG 358  
Qy 415 CTCAAGCTGGAAGAGTCCAGCTGAGCTTTGAGCGGATGGGAGTTGTCTTTGAACGT 474  
Db 359 TTAGCGCGAAGATGGCGGACTGT--TTTACAGATGATGCTGACTTCTGTTTAAAGC 415  
Qy 475 TTTGGGAGCGAGTCCAGAACTGATACCATCAACGACCTCGATTCAGGCGCATCTAT 534  
Db 416 TACGGCGATGCGTAAAGCACTGTTTACATTAATGAGCCAAAGATATGAGCGCTACTT 475  
Qy 535 GATATGCAACCGGCGAGCAAGCGCGCGGCGAGAGCATTTAACAGCACTCCACCGAG 594  
Db 476 GGTATGACACAGGTCMAATCTCTCTCAAGGTGCAACAGATGTG-----CTGCTGT 529  
Qy 595 GGAACAAGCTGCACTGAGCGCGTGGCTCGTGAAGGCGCCAGATCATGAGCCATGCCGC 654  
Db 530 GGAATTCAGCAACGGAACCTTATATGTTGCTCTATAATTTCTTGGGACATGCTACT 589  
Qy 655 GCCCTGGCGCTTACAGAGGAGCTTTGCGCCCTCGCAAAAGGCGCGAGTGGCATCTCG 714  
Db 590 GCAATTTCAAGATCCGTAACAAATATCAAGCTCTCAAGAGGTTAAGTGGGAATGTC 649  
Qy 715 CTCAAGCGGCACTACTATGAGCGCTGGGACAGCATGAGCTCGGAGCAAGAGAGGCT 774  
Db 650 CTGGACTTCACTGTAAGAGG---TCTTACAACTCACTGATGACCAAGAGAGAGCC 706  
Qy 775 GAGGAGCGGATGAATTTACATTTGGCTGTTGGCAATCCCATCTTTGAAGAAGAC 834  
Db 707 CAAGAGCGCAGGAGCTTCAATTTGGCTGTTGTTGATCC---AGCAGAGAGCGACAC 763  
Qy 835 TATCAAGAGACATGAAGAAGACAGCTGGGCGAGAGCTTCAAGCCCTCACTCCGCGGAC 894  
Db 764 TATCAACAGATTAATGAAGATCTCGTGAAGAGAGGCGCCAGAGTTCACTCTGAGAG 823  
Qy 895 TTGCGCATCTCAATGCGGAGAGACGACTTTCAAGCATGAATTAC---TACACATCC 951

Db 824 GCTAACTGTTGAAGGGCTCGGACGATCATCGGTATCAACAGATACACATCCAGCTAC 883  
Qy 952 CAGTTCGGCGGCGACCTTAAGAGGATCCCGTCCCGGAGCGGACATATCTCGGCGCATCAT 1011  
Db 884 ATGAAGGAGGAGAAAGTGTGTGCAAGTGGCGCGCGCGAGTGAATCTTGGCAGGTT 943  
Qy 1012 GAGCAGCAGAGATAAGACAGCGAGCGCGGTTGGCGAGAGAGAGCGGCTCGCTGGCTG 1071  
Db 944 CAATATGGCATGGCAGCCCAATGGCAACCATTTGACACAGCGGATTTCTAAGTGCCTC 1003  
Qy 1072 CGCTCTGCGCGGACATGTTCCGAAACATCTCGCCCGGATGATAGCGCTGTACGCGAAG 1131  
Db 1004 TACATCGCGCGGACGCGGAGTGTACGCGGTGGTGAATCACTCAAGAGAGATATGGGAAT 1063  
Qy 1132 C---CCATTCATCATACCGGAAACGATGCGCGCTCGCTGAGAGAGAAATGACCTGC 1188  
Db 1064 CCAACGATCTTACATTAAGAGAGAAACGAAATGAGACCAAGCTGGA-----AACTGACCGA 1117  
Qy 1189 GAGGAGCGCTCAACGACGCCCTTCCGATCCGATCTTGTGACTGCACTTGAAGCTGAT 1248  
Db 1118 GACCAATACCTGGCGGACGCGCACGAGGATCGGTTCTACAGAGCTATCGGCACTG 1177  
Qy 1249 TCCAGGCGATTAACCAAGAGCGCGCTGCTGCAAGGAGTACTTTGGTGGCGCTTCTC 1308  
Db 1178 AAGAGGCGCATTAAGCAGAGAGCGAAGCTGAC---TGGTACTTTCGCTGCTCTCTC 1234  
Qy 1309 GATTAATCTGAATATGATATGCTAGTACGACCCAGATTCGCGGCTGACGTTTACAGACTAC 1368  
Db 1235 GACAACTTCAAGTGGCTGAGAGGATCTGCTCAAGTTGCGCATCTGCTACGTGAGCTTC 1294  
Qy 1369 ACCACCTCAAGGCGACGCCCAAGAAAGTCTGCCCTGCTCTCAAGAGCATGTTT 1422  
Db 1295 AACACGCTCAAGCGCACCCGAAAGGCTCGGCTTACTGTTCAAGGACATGCTT 1348

RESULT 6  
CB677158 788 bp mRNA linear EST 09-APR-2003  
LOCUS  
DEFINITION OSJNBel13M03.f OSJNB Eryza sativa (japonica cultivar-group) cDNA  
ACCESSION CB677158  
VERSION CB677158.1 GI:29680883  
KEYWORDS EST.  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1 (bases 1 to 788)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Ming,R. and Wang,G.  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
Unpublished (2003)  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aac cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 13 row: N column: 03  
Seq primer: gta aac cga cgg cca gtc.  
Location/Qualifiers  
1..788  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Niponbare"

/db\_xref="taxon:39947"  
 /clone="OSJNE013N03"  
 /tissue\_type="leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNE"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)";

## ORIGIN

Query Match 13.4%; Score 195.2; DB 14; Length 786;  
 Best Local Similarity 56.3%; Pred. No. 2.5e-28;  
 Matches 409; Conservative 0; Mismatches 308; Indels 9; Gaps 2;

20 TGCACGAGACTTTTAAATGAGGCTTCGCAACGGCGCGCTACAGATGAGAGCGCGCTCA 79  
 60 TCCCGAGAGACTTATCTTCCGACCGGCTTTCAGTACAGAGAGCGCTTGA 119  
 80 AAGAGTGGCCCGCGCGCTTCATCTGAGACAGTACTGCACTGAGAGCGCTGCA 139  
 120 ACGAAGCGCGAGAGCGCGAGCATCTGGGACAGTACATTCAGGCAAGTTG 179  
 140 CCAAGCGCGCAAGCGCGAGTGGCTTCGATCACTACAGCGCTACAGATGAGACTTG 199  
 180 AAGATGTAGCAACGGGAGATGAGAGCTTTTACATGCTCAAGAGAGATCTGA 239  
 200 ATCTCTTACCAAGTACCGCGCAAGGCTTCCCTTCTCTGTCGTGCGGATCA 259  
 240 ACTTGTCGACCGACATGAAATGAGCGCTTCCGTTTCCATTCGCTTGAAGAGATCC 299  
 260 TTCCCTTGGCGGAGGCTGATCCGTCACAGAGAGAGAGATTTTACAGCAAC 319  
 300 TGCACATGAAACATCATGAGAGATCAAGAGAGAGATTTTCTTCAACAGCC 359  
 320 TGATTGAGCGCTTGTAGAGGCGGATACAGCGCTTGGGTGATTTGACACAGGATC 379  
 360 TGATTGAGAGATCATATCTAGAGGCTTGAAGCATTTGTACACATCTTCAATTGACA 419  
 380 TGCCTGAGCGCTTACAGATGCTATGAGAGCTGCTCAAGTGAAGAGTCCAGCTG 439  
 420 CCCACAGGCTTGGAGAGCAAAATACGAGCTTCTCA---GTGAAAACATTTGAAAG 476  
 440 ACTTTGAGCGGTATGCGAGGTTGCTTGAACGTTTGGGAGCCGAGTCCAGAACTGGA 499  
 477 ATTTCGTGACTATGCGAGCGTGTCTTCGTGAGTTGCGGAGCGGCTGAAGATGGA 536  
 500 TCACATCAACGAMCCCTGATTCAGGCTATGATGATGATGACCGCGAGCAAGCGCC 559  
 537 ACAAGTTCAAGAGCCATGATCTTTCGCGCGCGGCTACGCTTCGAGACCAAGGCCC 596  
 560 CGGAGAGAGCA-----CATTAACAAGCACTCCAGGAGGAGCACTGCACTGAGC 613  
 597 CCGGCGGCTGCTCCCGTACGCTTCAAGAAATGCGCGCGGCGAGCTCCGAGCAAGGAGC 656  
 614 CGTGGCTGCTGGAAGAGCCAGATCATGAGCCATGCCGCGCGCTGAGCGCTTACAGCA 673  
 657 CGTACGTGCGGCGCAACCTGCTGCTGCGCCAGCGGAGCGCTCCGCTTCAAGCGCC 716  
 674 GGAAGTTCGCGCGCTTCAAGAGAGGAGAGATCGGATCTGCTCAAGCGGAGCTACTATG 733  
 717 AAAAGTACAGAGCGAGAGAGAGGAGAGATCGGATCTGCAAGAGTGTGCACTGTTTCG 776  
 734 AGCCCT 739  
 777 TGCCCT 782

RESULT 7  
 LOCUS CB673495  
 DEFINITION OSJNE08A17.1 OSJNE Oryza sativa (japonica cultivar-group) cDNA  
 ACCESSION CB673495

## VERSION

CB673495.1 GI:29677220

## KEYWORDS

## SOURCE

## ORGANISM

Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Eubartoideae; Oryzaceae; Oryza.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

Antasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)  
 Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gtc  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 08 row: A column: 17  
 Seq primer: gta aaa cga cgg cca gtc.  
 Location/Qualifiers  
 1..786  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE08A17"  
 /tissue\_type="leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNE"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
 XhoI; 24 hrs after inoculation with Rice Blast (70-15)";

## ORIGIN

Query Match 13.3%; Score 193; DB 14; Length 786;  
 Best Local Similarity 57.1%; Pred. No. 6.9e-28;  
 Matches 437; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

66 CGAAGCGCGTAAAGAGTGGCGCGCGCTCATCTGGGAGACGATCCACCT 125  
 22 CAGGAGCTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81  
 126 GAGGAGATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185  
 82 CTTGGAAGATCAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141  
 186 CAGTGAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245  
 142 CAGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201  
 246 GTGTCGCGGATATTCCTCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 305  
 202 ATGTGTAAGATTCACCAATGAGT-----TTGATCAAGTCAATCAAGTGTGCA 255  
 306 GTTTACAGCAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365  
 256 CCATCAACAAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 315  
 366 GTACCACTGAGATCTGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
 316 CTACCACTGAGAGATCTGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 373  
 426 AGAGGTCCAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
 374 -CAGATGAGAGAGATTTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 432

QY 486 AGTCCAGAACTGATCATCAACATCAAGMCCCTGATTACAGCCATCTATGATATATGCCAC 545  
 Db 433 GGTGAAGACATGATGATCAAGCTCAACAGCGCACAGGTGGCCATCCAGGGCTACAGACGC 492  
 QY 546 CGGAGCAACAGCCCCGGGAGAGACAGATTA-----ACAGACCTCCACCGAGGCAA 599  
 Db 493 AGGGCTTCAGAGCCCCGGGCGCTCTCCGTGCTCCACCTCTACTGCAAGGCGGCAA 552  
 QY 600 CACTGCACATGAGCCCGGCTCGCTGGAAAGGCCAGATGATGAGCCATCCCGCGCGT 659  
 Db 553 CTCGGAGACGAGCCCTACGTCGTTCGCCACCACTTCATCTCCGCCACCGCGCGCGC 612  
 QY 660 GGCCTCTACAGAGAGGACTTTCGCCCTCGCAAAAGGCGCAGATCGGACTCTGCTCAA 719  
 Db 613 CACATCTTACAGAGCAAAATACAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 672  
 QY 720 CGGCACTACTATGAGCCTTGGAGACAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCG 779  
 Db 673 CTGATGTGCTGAGAGCGGATGTCACAA---CCACATGACATCGAGGCGGCAAGAG 729  
 QY 780 AGGATGGAATTCACATGCTGGCTGTTGCAATCCCATCTCTT 824  
 Db 730 AGCGCAGAGATTCACTAGAGATGTTGCTGATCCTTCTT 774

RESULT 8 CB642881 806 bp mRNA linear EST 08-APR-2003  
 LOCUS CB642881  
 DEFINITION OSJNB03E17.f OSJNB Oryza sativa (japonica cultivar-group) cDNA  
 ACCESSION OSJNB03E17.5', mRNA sequence.  
 VERSION CB642881  
 KEYWORDS CB642881.1 GI:29637872  
 SOURCE EST  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 806)  
 Jantschurikar,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
 Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 Large-scale identification of ESTs involved in the interaction  
 between rice and Magnaporthe grisea  
 Unpublished (2003)

## JOURNAL

COMMENT Contact: Rod Wing

Arizona Genomics Institute

University of Arizona

Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ

Tel: 520 626 3967

Fax: 520 621 9288

Email: http://genome.arizona.edu

PCR Primers

FORWARD: gta aac cga cgg cca gtc

BACKWARD: gga aac agc tat gac cat g

Plate: 03 row: E column: 17

Seq primer: gta aac cga cgg cca gtc.

Location/Qualifiers

1. 806

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="OSJNB03E17"

/issue\_type="Leaf"

/dev\_stage="3 week"

/lab\_host="DH10B"

/clone\_1b="OSJNB"

/note="Vector: pBluescript II KS +; Site\_1: EcoRI; Site\_2:  
 XhoI; 24 hrs after inoculation with Rice Blast (che  
 86061)"

Query Match 13.1%; Score 190.4; DB 14; Length 806;  
 Best Local Similarity 57.5%; Pred. No. 2,3e-27;  
 Matches 407; Conservative 0; Mismatches 286; Indels 15; Gaps 3;

QY 20 TGCCCAACGACTTTGAATGGGGCTTGGCAACGGCCCTTACAGATGAAAGGCGCGTCA 79  
 Db 96 TCCCAAGAGGAGTTCCTTCCTTCGAGCAGCGCTCCGCGCAGTACAGTACGAGGAGCTGTGA 155  
 QY 80 AAGAAGTGGCGCGGCGCGCTCCATCTGGGAGACGTAAGTCCGACCTGAGACCTCGGCA 139  
 Db 156 AGGAGAGCGGAGAGGAGGAGACCATCTGGAGACGTTCCGACACCTTTGGAAAGATCA 215  
 QY 140 CCAAGCGCGCAACGCGAGATGTGGCTTGGCATCTACACCGCTACGATGAGAGACTTTG 199  
 Db 216 CGACATTCAGCAATGCTGATGTTGAGATTGATCAGTACACCGTTTCGAGGAGATATAC 275  
 QY 200 ATTCTTACCAAGTACGCGGCAAGGCGCTACCGCTTCTCTTGTGTGTGCTCGGATCA 259  
 Db 276 AGCTCATGCGAGACATAGGAGATGATGCGTATGCTTCGATAGCATGATGTCAGATCT 335  
 QY 260 TTCCTCGCGCGGCGAGCTGATCCGCTCAACGAGAGGAGAAATTGATTTACAGCAAC 319  
 Db 336 ACCCAATAGTG-----TTGATCAAGTCAATCAAGCTGATGACACTACACCAAGC 389  
 QY 320 TGATGACGCGCTGTGAGCGGCGGTATACAGCTTGGGTGACTTGTACACTGCGGATC 379  
 Db 390 TGATGATGACTCTTAGCAAAAGAAATTGACGATATGACACTTACACTGCGGAGC 449  
 QY 380 TGCTTACGCGGCTTACGATCGTATGAGAGCTGCTCAACGTTGAAGAGGTCCAGCTGG 439  
 Db 450 TCCCGCAGGCGCTTGAAGACAACTACAGGCGCTGCTTACAGAG---CAGATGTGAGAG 506  
 QY 440 ACTTGAAGGCGTATGAGAGGTTGTGCTTGAAGCTTTGGGAGCGAGTCCAGAACTGA 499  
 Db 507 ATTTCCGCGGCTAGCGGAGACCTGTTCAGGAGTTCCGAGAGGAGGTACACACTGA 566  
 QY 500 TCAACATCAACGAGMCCCTGATTCAGGCACTATGATGATGACACCGGAGCAAGCGCC 559  
 Db 567 TCAAGCTCAACGAGCGGACACAGCTGCGCATCCAGGAGCTACGAGAGGCTCAAGGCC 626  
 QY 560 CGGCGAGAGAGCAGC-----ATTACAGACTCCACCGAGGCAACACTGCGACTGAGC 613  
 Db 627 CCGCGCGCTGCTCGGTGCTCCACCTCACTGCAAGCGCGCAACTCCGCGACAGC 686  
 QY 614 CGTGCCTCGCTGAAGAGGCCAGATGATGAGCCATCCCGCGCGTGGCTTACAGCA 673  
 Db 687 CTAAGCTGCTGCGCCACCACTTCATCTCGCCACCGCGCGCGCAGCATCTACAGGA 746  
 QY 674 GGGACTTTCGCGCCCTCGCAAAAGGCGCAGATCGGACTCTGCTCAAG 721  
 Db 747 CAATAATCAAGGAGCGACGCAAGCGGCGACTTGGATAGCTTCAGAG 794

RESULT 9 CD438244 889 bp mRNA linear EST 03-JUN-2003  
 LOCUS CD438244  
 DEFINITION E01JN0511A05.b Endosperm\_5 Zea mays cDNA, mRNA sequence.  
 ACCESSION CD438244  
 VERSION CD438244.1 GI:31353887  
 KEYWORDS EST.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 889)  
 Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and  
 Messing,J.  
 Sequencing of the maize endosperm ESTs  
 Unpublished (2002)  
 CONTACT: Lai, Jinheng  
 DR. Joachim Messing's lab  
 Wakeman Institute, Rutgers University

ORIGIN

190 Freilingshuyzen Rd., Piscataway, NJ 08854, USA  
Tel: 732-445-3801  
Fax: 732-445-5735  
Email: jla@wakeman.rutgers.edu

Seq primer: T3

Location/Qualifiers

1. .889  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultiivar="W22"  
/db\_xref="taxon:4577"  
/cruise\_type="Endospem\_5"  
/clone\_lib="Endospem\_5"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

# ORIGIN

Query Match 12.3%; Score 179.2; DB 14; Length 889;  
Best Local Similarity 54.8%; Pred. No. 4e-25; Indels 24; Gaps 6;  
Matches 493; Conservative 0; Mismatches 383

52 GCCGCTACCAAGTGAAGGCGCGTCAAGAGGTGGCGCGCCGCTCCATCTGGGAC 111  
5 GCGGTGACAGAGTGAAGGCGCGCGTCAAGAGGCGCGCGCTCCATCTGGGAT 64  
112 ACGTACCTGACCTGGAGCCATCGGCAACAGCGCGCCAGCGGATGGCTTGGAT 171  
65 TCATTGCGGCAAGTCCAGAAATTTGACAGGAATCAAAATGAGAGCTTGGAT 124  
172 CACTACCAACCGTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCAAGGCTTAC 231  
125 CAATACCACTGCTACAGAGAAAGCTCATCTCATGAAAGTTGAATTTGATGCTTAC 184  
232 CGCTTCTCCTTGTGCTGTGGGATCAATTCCTCGCGGCGAGGCTGGATCCGTCAC 291  
185 CGTTCTCAATCTCATGCTGCGAGATCTCCGAGTGGCGAAGGG-----AAAGTCAT 238  
292 GAGGAGGAATGATTTTACAGCAAACTGATGAGCGCGCTGTGAGGCGGGGTATCAG 351  
239 CCAGAGGAGTGAAGCTATTAATTAATTTGATTAATCTATCTTTCAGCAAGGCTAGCT 298  
352 CCTTGGGATGACTTTGATCACTGGATGCTGCTCAGCGGCTTCAAGATCGCTATGAGAGC 411  
299 CCTTACATCAACTTTACCATATGATCTTCTCTTGGCGCTTGAAGAAATATGAGAGG 358  
412 TGGCTCAACGTGAAGAGTCCAGCTGACCTTTGAGCGGTATGCGAGGTGCTTGA 471  
359 TGGTAAAGCGCAAGATGCGGAGCTTG---TTTACAGACTATGACTTCTGTTTAAAG 415  
472 CGTTTGGGAGCGGAGTCCAGAACTGGATCACCATCAAGAGCGCTGATTCAGGCCATC 531  
416 ACCTACGAGGATGCGTAAAGCATGCTGTTTACATTAAGAGCCAAAGATATAGCGCTA 475  
532 TATGATATGCAACCGGAGCAACCGCCCGGAGAGGAGAGCATTAACAGCACTCCACC 591  
476 CTGGCTATGACACAGGATGCTCAATCTCTCAAAAGGTG-----ACCAATGCGGTGCT 529  
592 GAGGCAACATCTCCCTGAGCGGTGCTGCTGGAAGGCCAAGTCAATGAGCATGCG 651  
530 GGTGGGAATTCAGCAACCGAACTTACATAGTGTCTCAATATTTCTCTTGGCAATGCT 589  
652 CGGCGGTGGCGCTTACAGCAGGAGCTTGGCGCTCCGCAAAAGGCGAGATCGGATC 711  
590 ACTGCAAGTTGCAAGATACGTAAGAAATATCAGGCTGCTCAGAGGATTAAGGTCGGAATA 649  
712 TCCTCAACGCGCACTATATGAGCCCTTGGAGAGCAATAGGCTTCGGAGCAAGAGCT 771  
650 GTCTGGAATTTCACTGTATGAGAGGCTTT---ACAAATCACTATGATGCAAGAGCA 706  
772 GCTGAGAGGAGGATGGAATTTCAATGCTGTTGGTTCATCCATCTTCTTGAAGAG 831  
707 GCCCAAGAGCCAGGAGCTTCAATTTGCTGTGTTGATCC---ATTGATTAACGGA 763

QY 832 GACTATCCAGAGCATGAAGAAGAGCTGGCGAGAGGCTTCCAGCCCTCACTCCCGG 891  
DB 764 CACTATCCAGATATATGAGAAATCTCGTGAAGAGGCTGCCAGGTTCACTCCTGAG 823  
QY 892 GACTTGGCATCTTCATGCTCCGAGAGACCACTTCTAGGCAATGATTTACATCAATCC 951  
DB 824 CA---GGCTAACTGGTGAAGGCTTCGCGCACTACATGATGATTAACAGATCAATCC 880

RESULT 10  
CB688082 480 bp mRNA linear EST 04-SRP-2003  
LOCUS CEST-23-E-06 Mixed source, strain EP15 and EP15 infected with  
DEFINITION hypovirus CHV1-EP13 Cryphonectria parasitica cDNA clone EP15,  
EP15-CHV1-EP13 5-prime, mRNA sequence.

ACCESSION CB688082  
VERSION CB688082.1 GI:34448951  
KEYWORDS EST.  
SOURCE Cryphonectria parasitica  
ORGANISM Cryphonectria parasitica  
Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetidae; Diaporthales; Valsaceae; Cryphonectria-Endothia  
complex; Cryphonectria.

REFERENCE 1 (bases 1 to 480)  
Dawe, A.L., McMaine, V.C., Panglao, M., Kasahara, S., Chen, B. and  
Nuss, D.L.  
An ordered collection of expressed sequences from Cryphonectria  
parasitica and evidence of genomic microsynteny with Neurospora  
Microbiology 149 (9), 2373-2384 (2003)

TITLE JOURNAL  
MEDLINE 22830414  
COMMENT Contact: Dawe AL, Nuss DL  
Center for Biosystems Research  
University of Maryland Biotechnology Institute  
5115 Plant Sciences Building, College Park, MD 20742, USA  
Tel: 301-405-7661 or 5111  
Fax: 301-314-9075  
Email: dawe@umd.umd.edu  
Cryphonectria parasitica EST derived from mixed samples of cDNA  
from wild-type (strain EP15) and hypovirus-infected cultures.  
Seq primer: M13-reverse.

FEATURES  
source  
Location/Qualifiers  
1. .480  
/organism="Cryphonectria parasitica"  
/mol\_type="mRNA"  
/strain="EP15"  
/db\_xref="taxon:38755 and 52571"  
/db\_xref="taxon:5116"  
/clone\_lib="Mixed source, strain EP15 and EP15 infected  
with hypovirus CHV1-EP13"  
/note="Cryphonectria parasitica EST derived from mixed  
samples of cDNA from wild-type (strain EP15) and  
hypovirus-infected cultures"

# ORIGIN

Query Match 12.1%; Score 175.6; DB 14; Length 480;  
Best Local Similarity 64.4%; Pred. No. 1.6e-24;  
Matches 290; Conservative 0; Mismatches 158; Indels 2; Gaps 2;

QY 11 CGTAGCTTGCCTCAACAGCACTTTGAATGGGCTTCGCAACGCGCGCTACAGATCGAAG 70  
DB 4 CGCGTCCGCTACATCAACTTCAAGTGGGCTTTGCGACGCGCTCTCAAGGTCGAAG 63  
QY 71 GCGCGCTCAAGAGGATGGCGCGCGCTCATCTGGGACAGCTACTGCACTCGAGC 130  
DB 64 GCTTACCGACAAGAGCGCGCGCGCTCATCTGGGACAGCTTCTGAGATGCCG 123  
QY 131 CATCGGACCAACGCGCGCAACGCGGATGGCTTGCATCACTACCAAGGCTACATG 190  
DB 124 GCAATATGCGGAGCGGCTCCAGCGGCGGTGGCTGCAAGCTACCAAGCGCGGCG 183



OY	AGGACCTTGGATCTCTTGGACCAAGTACGGGGCAAAAGCCCTACCGCTTCTCTTGATGGT	250
Db	AGGACATTGCGCTGCTCAAGACGCTCGGGGGCGACGGCGTCTTCGCTGTCTGGT	243
OY	CGCGGATCATTTCCCTCGGGGGCAGCGTGGATCCCTCAACGAGAGGAAATTGATTTT	310
Db	CGCGCATATCCCGCTGGGGGGGGCCGACGACCCCGTCAACCAAAAGGGGCTGGACCACT	303
OY	ACAGCAAACTGATTGACGCCCTGTTGAGGGCGGGGATACAGCGCTTGGGATTTTGACT	370
Db	ACAGGCGCTTACCAACGACCTGCTGGGGGGCGGGATCAGAGCCCTTCATCAGCGTGTACC	363
OY	ACTGGATCTGCTCCAGAGCGCTTACGATGCTATGAGG-CTGACTCAACGTGGAAGA	428
Db	AACTGGGACCTGCCCGACGAGANTGGACCGGCGGTACGGGGGCTCTGTAAACCGACCA	423
OY	GGTCCAGCTGACCTTGAAGCGGTAGCGAG	458
Db	GTTCGGCTCGACTTGGCCGGGTACGGGG	453

RESULT 11			
AY106991			
LOCUS	AY106991	1182 bp	mRNA
DEFINITION	Zea mays PC0088410 mRNA sequence.		linear
ACCESSION	AY106991		
VERSION	AY106991.1	GI:21210069	
KEYWORDS	HTC.		
SOURCE	Zea mays		
ORGANISM	Zea mays		

REFERENCE	1 (bases 1 to 1182)
REFERENCE	Hanney, C. F., Dolan, M., Miao, G. H., Vogel, J. M., Whitsett, M. S.,
AUTHORS	Arthur, L. W., Hanafey, M., Morgance, M., and Tingey, S. V.
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of
JOURNAL	Overgo Probes
REFERENCE	Unpublished (2002)
REFERENCE	2 (bases 1 to 1182)

**JOURNAL COMMENT**  
Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA  
If you are interested in getting corresponding physical clones, these are publicly available from zmbd and may be found by BLAST searching at [MSL, maizemap.org](http://MSL.maizemap.org); [ZmDB, www.zmdb.iastate.edu](http://ZmDB.zmdb.iastate.edu); [TIGR, www.tigr.org](http://TIGR.tigr.org); or [NCBI, www.ncbi.nlm.nih.gov](http://NCBI.ncbi.nlm.nih.gov). When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from [ZmDB, www.zmdb.iastate.edu](http://ZmDB.zmdb.iastate.edu).

```

FEATURES
source
location/Qualifiers
1. 1182
/organism="Zea mays"
/mol_type="mrna"
/db_xref="MaizeDB:635002"
/db_xref="taxon:4577"
/clone_11b="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

```

Query March	12.0%	Score 175;	DB 11;	Length 1182;
Best Local Similarity	54.8%	Pred. No. 3.1e-24;		
Matches 440;	Conservative	0;	Mismatches 345;	Indels 18;
			Gaps	4;
0y	16	GCTCTGCCCAAGACTTGTGAATGGGGCTTTCGCAAGCGCCGCTTACCACAGTCGAAAGGCC	75	

Db	205	GCGTTTCCCGGACGGGCTTGCTCTTGCGGAGCGGCGCGGTACGGGATACAGAGTGGAGGGAGT	264
Oy	76	GTCAAAGAGGTGGCGCGCGCCCGCTCCATCTGGAGACAGCTACGTGCACCTGGAGCCATCG	135
Db	265	GCCAAAGCAGCGCGGGCGGGCCCGACATCTGGGACGGCTTCACTAGAGGTTCCGGGACC	324
Oy	136	CGCACCAACGGCGCCAACGGCGGATGTGGCTTGGCATCACTACACCGCTACGATGAGAAC	195
Db	325	ATCCCTTAACAAATGCCACCGCTGACGTGACGGTGCAGGTATCACTCGGTACCAAGAGAT	384
Oy	196	TTTGATCTCTTGAACCAAGTACGGGCGCAAGCGCTACCGCTTCTCCTTGTCGATGTCGGG	255
Db	385	GTGAACATTAAGAAAGAACATGAGCGCTTGATGCGTACCGGTTTGCATCTCTTGGTCAGG	444
Oy	256	ATCATTTCCCTTCGGCGGCGAGCTGGATCCGCTCAACGAGAGAGGAATTGAGTTTACAGC	315
Db	445	ATTTTCCAGATGGAACCTGGCAG-----GTGAACCAAGAAAGAGTGGATTACTACAAAC	498
Oy	316	AAACTGATTGAACGCCCTGTGTGAGGGCGGGGTATCAAGCCTTGGGTGACTTTGTACCACTGG	375
Db	499	AGGCTCATAGATTACATGCTCCAGAAAGTATCGGCGGCTATGCAAACTCTTACCATTAAT	558
Oy	376	GATCTGCTTACAGCGGCTTACAGATCGCTATGAGAGCTGAGCTTCAACGTGGAAGAGGTCCAG	435
Db	559	GACCTCCATTGGCACTCATGTAACAGTACCTGGGCTGGCTTAAGCC--CAAGATTGG	615
Oy	436	CTGCACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAAC	495
Db	616	GAGGGTTTGGAGACTACGCGAGTTCGTCTTCCACGGCTTCGGAGAACAGGGTGAAGAAC	675
Oy	496	TGATACACATCAACGAACCTCTGATTCAGGCACTATATGATATATGACACGGGACGAAC	555
Db	676	TGTTTACCTTCAACGAGCCGAGGTGCTGCTGCTCTGGGCTACGACAAATGGCTTGCAC	735
Oy	556	GCCCCGGGACAGACAGCATTTAAACAGCACTTCCACCGAGGGCAACACTGCCACTGAGCCG	615
Db	736	GCACCGGGAAGTG-----TTCCGGGTGCCCGCGGAGGCAACTCCACACAGGAGCCG	789
Oy	616	TGGCTCGCTGAAAAGGCCAGATCATGAGCCATGACCCGCGCGTGGCCGTCTACAGCAGG	675
Db	790	TACCTGTGCGACACATCTCATCTTCTCATGAGGTGGGGTGAAGCATACCGGCAC	849
Oy	676	GACTTTGCCCCCTCGCAAAAGGGCCAGATCGGACATCTCGTCAACGGGCACTATATGAG	735
Db	850	AAGTATACGTTTCAACGAAAGGGGAAGATTGGAATTCTCTTGATTTTGTTGTGTACGAA	909
Oy	736	CCCTGGGACAGCAATGAGCTTCGGGACAAAGAGGTGCTGAGCCACGAGATGGAATTTTAC	795
Db	910	CCCTT---TCAGCGACAGCAATGCGGACCGAGCTGACAGCAGCGAGCCAGGGAATTTCCAC	966
Oy	796	ATTGAGCTGTTTGGCAATCCCAT	818
Db	967	CTAAGCTGTCTTACCCCAT	989

RESULT 12				
LOCUS	CF880605			
DEFINITION	CF880605	662 bp	mRNA	linear EST 31-OCT-2003
ACCESSION	tricc082xh20.d1.T.reesei	mycelial culture		Version 6 October 2003
VERSION	CF880605			
KEYWORDS	CF880605.1	GI:38135287		
SOURCE	EST.			
ORGANISM	Hypocrea jecorina (anamorph: Trichoderma reesei)			
	Hypocrea jecorina			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
REFERENCE	Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.			
AUTHORS	1 (bases 1 to 662)			
	Diene, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,			
	Mitchell, T.K., van Solingen, P., Teunissen, F.J.M., Ward, M., and			
	Dean, R.A.			
TITLE	Analysis of the protein processing and secretion pathways in a			
	Trichoderma reesei EST dataset			

## JOURNAL

Unpublished (2003)

CONTACT: Ralph A. Dean  
Fungal Genomics Laboratory  
North Carolina State University  
Campus Box 7251, Raleigh, NC 27695, USA  
Tel: 919-513-0020  
Fax: 919-513-0024  
Email: ralph.dean@ncsu.edu

Seq primer: LT-F1 primer.  
Location/Qualifiers

## FEATURES

source

1. .662  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="OM6a"  
/db\_xref="taxon:51453"  
/clone="tric082x120"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 6 October 2003"  
/note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## ORIGIN

Query Match 11.8%; Score 171.8; DB 14; Length 662;  
Best Local Similarity 59.4%; Pred. No. 1e-23;  
Matches 366; Conservative 0; Mismatches 238; Indels 12; Gaps 4;

QY 585 CTCACCGAGGCGCAACTGCGATGAGCCGTGCTGCTGGAGAAAGGCCAGATCATGAG 644  
DB 52 CGCCCCCGCGCGAGACACCTCGAGCCGTGGACCTCGGCCCAACATCTCTGTGCG 111  
QY 645 CCATGCCCCCGCGCGCGCTCTACAGAGGAGCT---TGCCCCCTCGCAAAAGGGCCA 701  
DB 112 CCAGCGCGCGCGCGCGAGCGGTACCGAGACTTCAAGCCCGCCAGCGCGAGCGCA 171  
QY 702 GATCGGCACTCTGCTCAACCGCGACTACTATGAGCCCTGGAGACAGCAATGAGCCCTGGGA 761  
DB 172 GATCGGCACTCTGCTCAACCGCGACTTCAACCTCGGGAGCGCGCGAGCCCGCGCA 231  
QY 762 CAAGAGGCTGTGAGCGAGCGAGTGAATTTCACATTGCTGTGTTGCCAATCCATCTT 821  
DB 232 CAAGAGGCGCGCGCGCGCGCTCGAGTTCTTCAAGGCTGTGTTGCCAATCCATCTT 291  
QY 822 CTGGAAGAAGACTTCCAGAGAGCATGAAAGAGCTGGCGAGAGGCTTCCAGCCCT 881  
DB 292 CTT---GGGCGACTACCGGCGCTCGATGCGAGAGAGCTGGCGAGCCGCTGCCACTT 348  
QY 882 CACTCCCGCGAGCTTGGCATCTCTCAATGCCGAGAGACCGACTTCTACGGGATGAATTA 941  
DB 349 TACGCCGAGAGCGCGCGCGCTCTGTCCA---CGGCTCCAAAGACTTTTACGGGATGAACCA 405  
QY 942 CTACACATCTCCAGTTGCGCGCGCTCTAGAGCGGTCCCGTCCCGAGAGAGATATCTGG 1001  
DB 406 CTACACATCTCCAGTTGCGCGCGCTCTAGAGCGGTCCCGTCCCGAGAGAGATCTGG 465  
QY 1002 CGCCATTCATGAGCAACAGAGATTAAGAGCGAGCGCCCGTTGGCGAGAGAGCGGCT 1061  
DB 466 CAAGCTGCGAGCTGCTTTCACCAAGAGAGGCAATGATCGGCGCGAGAGCGGATG 525  
QY 526 CCCCTGGTGGCGCGCGCGCTGTGGCGGCTTCCGCACTTCTGTGTGATCAGACAGAG 585  
DB 1122 GTACGGCA---GCCATCTAATCAACGAGAAAGGATCCGTCCTCGAGAGAGAGA 1178  
QY 586 GTACGGCAATCCGCGCATCTACGTAGCGAAGAGCGACGAGCATTAAGAGGAGAGCGCA 645  
QY 1179 CATGACGTGAGAGAG 1194  
DB 646 CTTTGCCCAAGAGAGAG 661

## RESULT 13

CB907493

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES  
source

Contact: Pamela K. Foreman  
Genencor Intl.  
925 Page Mill Road, Palo Alto, CA 94304, USA  
Tel: (650) 846-7635  
Fax: (650) 621-7817  
Email: pforeman@genencor.com  
Seq primer: LT-F1 primer.  
Location/Qualifiers

1. .752  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="OM6a"  
/db\_xref="taxon:51453"  
/clone="tric082x120"  
/dev\_stage="mycelia"  
/clone\_lib="T.reesei mycelial culture, Version 3 april"  
/note="Vector: PREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

## ORIGIN

Query Match 11.8%; Score 171.8; DB 14; Length 752;  
Best Local Similarity 59.4%; Pred. No. 1.e-23;  
Matches 366; Conservative 0; Mismatches 238; Indels 12; Gaps 4;

QY 585 CTCACCGAGGCGCAACTGCGATGAGCCGTGCTGCTGGAGAAAGGCCAGATCATGAG 644  
DB 142 CGCCCCCGCGCGAGACACTTGGAGCGGTGAGACGTGGCCCAACATCTCTGTGCG 201  
QY 645 CCATGCCCCCGCGCGCGCTCTACAGAGGAGCT---TGCCCCCTCGCAAAAGGGCCA 701  
DB 202 CCAGCGCGCGCGCGCGAGGTACCGGAGACTTCAAGCCCGCCAGCGCGAGCGCA 261  
QY 702 GATCGGCACTCTGCTCAACCGCGACTATGAGCCCTGGAGACAGCAATGAGCCCTGGGA 761  
DB 262 GATCGGCACTCTGCTCAACCGCGACTTCACTTACCTCTGGAGACCGCGAGCCGCGCA 321  
QY 762 CAAGAGGCTGTGAGCGAGGAGATTAATTCATGAGCTGTGTTGCCAATCCATCTT 821  
DB 322 CAAGAGGCGCGCGAGCGCGCTCGAGTTTTCACGCGCTGTGTTGCCGATCCATCTT 381  
QY 382 CTT---GGGCGACTACCGGCGGTGATGCGAGAGAGCTGGGCAACGCGGCTGCGACTT 438  
DB 822 CTGGAAGAAGACTTCCAGAGAGCATGAAAGAGCTGGGCGAGAGGCTTCCAGCCCT 881  
QY 382 CTT---GGGCGACTACCGGCGGTGATGCGAGAGAGCTGGGCAACGCGGCTGCGACTT 438  
DB 882 CACTCCCGCGAGCTTGGCATCTCTCATGCGGAGAGACCGACTTCTACGGGATGAATTA 941  
QY 439 TACGCCGAGAGAGCGCGCGCTGTGCTCA---CGGCTCCAAAGCACTTTTACGGGATGAACCA 495  
DB 942 CTACACATCTCCAGTTGCGGCGCACTTGAAGAGTCCGCTCCCGAGAGCGACTATCTCGG 1001

Db 496 CTACAGCTCCACTATATCCGCCACCGCAGCTGCGCCCTCCGCGCAGACACCTGTGG 555  
Qy 1002 CGCATCATAGAGACAGGAGATTAAGAGCGGAGCGCCGTTGGGAGAGAGCGGCTT 1061  
Db 556 CAACGCTGAGCTCTTTACCAACAGCAGGAGAACTGATCGGCCCGCAGACGAGTC 615  
Qy 1062 CGCTGAGCTGCGCTCTGCGCCGAGCATGTTCGGAAGCATCTGCGCCGCTGTACGCGCT 1121  
Db 616 CCCCTGAGCTGCGCCCTGTGCGCCGCGCTTCCGGAACCTTCGTGTGTGATCAGNCAGAG 675  
Qy 1122 GTACGCGCA---GCCCATCTACATCACCGAAGACGATGCCCGCTTGGAGAGAGAA 1178  
Db 676 GTACGCGCTACCCCGCATCTACGTGACGGAACGCGACGAGCATCAAGGCGAGAGCGA 735  
Qy 1179 CATGACGTGAGAGAG 1194  
Db 736 CTTGCCCAAGAGAGAG 751

RESULT 14  
BE777259 650 bp mRNA linear EST 20-SEP-2000  
LOCUS BE777259  
DEFINITION MY-26-C-10 PinfeestansMY Phycophthora infestans cDNA, mRNA sequence.  
ACCESSION BE777259  
VERSION BE777259.1 GI:10230914  
KEYWORDS EST.  
SOURCE Phycophthora infestans (potato late blight agent)  
ORGANISM Phycophthora infestans  
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;  
Phytophthora.  
1 (bases 1 to 650)  
Kamoun, S., Hraber, P.T., Sobral, B.W.S., Nuss, D. and Govers, F.  
Initial assessment of gene diversity for the oomycete pathogen  
Phycophthora infestans based on expressed sequences  
Fungal Genet. Biol. 28 (2), 94-106 (1999)  
20056376  
10587472  
Contact: Govers F  
Laboratory of Phytopathology  
Wageningen University  
Bilthoven 9, P.O.Box 8025, 6700 EE, Wageningen, The Netherlands  
Tel: 31 317 483 138  
Fax: 31 317 483 412  
Email: Francine.Govers@medew.fyco.wau.nl.

FEATURES  
source  
1.650  
Location/Qualifiers  
/organism="Phycophthora infestans"  
/mol\_type="mRNA"  
/strain="DDR7602, A1 mating type"  
/db\_xref="taxon:4787"  
/dev\_stage="4-week old vegetative, non-sporulating  
mycelium in synthetic medium"  
/lab\_host="E. coli, strain DH5-alpha"  
/clone\_idb="PinfeestansMY"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; Total  
RNA was isolated from mycelium of P. infestans DDR7602  
cultured for 4 weeks in synthetic medium. EST clones were  
named by their position in the microtiter plate, preceded  
by the prefix MY (for mycelial) and the successive number  
of the microtiter plate (e.g. MY-06-A-04)."

ORIGIN  
Query Match 11.7%; Score 169.8; DB 10; Length 650;  
Best Local Similarity 58.1%; Pred. No. 2.6e-23;  
Matches 319; Conservative 0; Mismatches 224; Indels 6; Gaps 1;

Qy 20 TGCCCAACGACTTGTATGAGGCTTCGCAACGCGCTACCAATGAGGCGCGCTCA 79  
Db 58 TCCCGAAGAGCTTATGTGGGACATGCCCTCTGTACCAAGTGGAGGCTGACGA 117  
Qy 80 AAGAAGGTGGCGCGCGCTTCATCTGGGACACGTAATCCACCTTGAGACCATCGGCA 139  
Db 118 ACGAAGGTGAGCGCGGACAGCATCTGGACGCTTCTCTCGGACACCGGCAAGATCG 177

Qy 140 CCAACGCGCAACGCGGATGTGCTTGGCATCATCACCGCTACGATGAGGACTTTG 199  
Db 178 TGAATGAGACACGCGGCGAGAAAGCCGTGATATATATCTATCCCTCAAGAGAGACTGC 237  
Qy 200 ATCTTTGACCAAGTACGCGGCAAGGCGCTTACCGCTTCTCTTGTGTGCGGATCA 259  
Db 238 AACTTATGAAGAAGATGGGACTCAAGACTTACCGCTCTTCATCGCTTGGCCCGGATCA 297  
Qy 260 TTCCCTCGCGGAGGCTGTGATCCCGTCAACGAGAGGAAATTGATTTACAGCAAC 319  
Db 298 TCCCGCAGAG---TCGCGCAGTGAACGAGAGCGCTGAGTTTCAACAAC 351  
Qy 320 TGATTAACGCGCTTGTGAGGCGGCGGATATACAGCGCTTGGGATCTTTGATCACTGGATC 379  
Db 352 TCATCAATGAGCTACTGCGCCACGACATACAGCCCTTGATAGCTGTACATTTGGATT 411  
Qy 380 TGCTCAGGCGCTTACGATCGCTATGAGAGCTGTCAACGTGAGAGAGTCCAGCTGG 439  
Db 412 TACCCTTAGCTTGCAACCGAGTATGACGGTTGGCTTGGTGGCAGATCATCAAGACG 471  
Qy 440 ACTTGAAGCGGTATGCGAGTGTGCTTGAACGTTTGGGACCGAGTCCAGACTGGA 499  
Db 472 CTTTGTGCAATACGCTCGGTGTGCTTCCAGGCTTCGAGACCGTGTGACCAACTGT 531  
Qy 500 TCACATCAACGACGCCCTGATTCAGGCCATCTATGATATGCGACCGGACGACGCC 559  
Db 532 TGACGTTGAATGAGCCGTGTGTGCTCCGCTTTTGTGGCTATGGAATGCTTCAATGCTN 591  
Qy 560 CGGCGAGGA 568  
Db 592 CAGACGCA 600

RESULT 15  
AJ436490 613 bp mRNA linear EST 15-MAR-2002  
LOCUS AJ436490  
DEFINITION AJ436490 S00007 Hordeum vulgare cDNA clone S000070006A08F2, mRNA  
sequence.  
ACCESSION AJ436490  
VERSION AJ436490.1 GI:19524942  
KEYWORDS EST.  
SOURCE Hordeum vulgare  
ORGANISM Hordeum vulgare  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Hordeum.  
1 (bases 1 to 613)  
Saren, A.-M., Tanskanen, J., Paulin, L. and Schulman, A.H.  
Barley EST's  
Unpublished (2002)  
Contact: Schulman AH  
Institute of Biotechnology  
University of Helsinki  
P.O.Box 56 (Viikinkaari 6A), University of Helsinki FIN-00014,  
Finland.

FEATURES  
source  
1.613  
Location/Qualifiers  
/organism="Hordeum vulgare"  
/mol\_type="mRNA"  
/db\_xref="taxon:4513"  
/clone="S000070006A08F2"  
/dev\_stage="shoot"  
/clone\_idb="S00007"  
/note="2,-3,-4-days after germination"

ORIGIN  
Query Match 11.2%; Score 162.6; DB 9; Length 613;  
Best Local Similarity 57.2%; Pred. No. 6.7e-22;  
Matches 338; Conservative 0; Mismatches 244; Indels 9; Gaps 2;

Qy 143 ACGGCGCAACGCGGATGTGCTTGGCATCATCAACGCGCTTACGATGAGACTTTGATC 202

```
Db      19  ACGGCAGCAAGTGTATACAGAAACGACTTCTATCATCCGCTACAAGAGGATCTGAAGC 78
QY      203  TCTTGACCAAGTACGAGGGCAAAGGCGCTACCGGCTTCCTGTCGTGTGCGGATCATTC 262
Db      79  TGGTAAACGATGAATGAGCGCTTCGCGTTCCTCTTGATGAGAGGATCCTTC 138
QY      263  CCTCGGCGGAGGCTGATCCGCTCAACGAGAGGGAATTGAATTTTACAGCAAACTGA 322
Db      139  CAATGGAACCATGCGCGGAGGATCAACAAGCAGAGTGAATTTCTACACAGCCTCA 198
QY      323  TTGACGCCCTGTGAGCGGCGGTATCAAGCCTTGGGTGACTTGTACACTGGGATCTGC 382
Db      199  TCGACGAGGTTTGGCTAAGAGGCTGATGCTTTCGTACAAATGTTCCACTTGACACCC 258
QY      383  CTCAGGCGCTTCAAGATCGCTATGAGGCTGGCTCAACGTGAAAGAGTCCAGCTGACT 442
Db      259  CCCAGGCCCTCGAGGACAAATACGGAAGCTTCTTGAGC---GACACATCGTMAAGGACT 315
QY      443  TTGACCGGTATGCGAGGTGTGCTTTGAAAGTTTGGGGAACGAGTCCAGAACTGATCA 502
Db      316  ACGTGAGTACGCGGAGCTGTGCTTCAAGCTGTTGCGGACCGGAGTGAATTTGACCA 375
QY      503  CCATTAACGAMCCCTGGAATTGAGGCATCTATGATATGACCAGGAGCAACGCCCGG 562
Db      376  CTTTAACGAGCCATGAGTGTCTGCGCTTCGCGCTACGGAACCGGACCTTGGGCCCGG 435
QY      563  GCAGGAGC-----AGCATTTAACAAAGCACTCCACCGAGGCAACACTGCCACTGAGCCGT 616
Db      436  GCCGCTGGCTCGGTAACGTCTCCAAAGGCTGGCGGCGCCGGGAGCTCTCCACGAGGCCCT 495
QY      617  GGCTGCGTGGAAAGGCCCAATCATGAGGCAATGCCGCGCGCGTGGCCGCTTACAGCAGG 676
Db      496  ACATGCGCGGCAACCTCTCATGCCCACAGCCGAGGCGGTGCACTGTATCCGACCA 555
QY      677  ACTTTCGCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGGACT 727
Db      556  GGTAACGCGCGGCGAGCGGCGGCGGAGATGGCATCGTGCAATATCCCACT 606
```

Search completed: March 28, 2004, 23:31:39  
Job time : 3907.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 17:03:34 ; Search time 5752.87 Seconds  
(without alignments)  
10962.209 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455  
Sequence: 1 atgcccagctcctagctctc.....gggttaagtgccgcatcaa 1455Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

GenEmbl: \*  
1: gb ba: \*  
2: gb hcg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_stb: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_ov: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_stb: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hcg\_hum: \*  
31: em\_hcg\_inv: \*  
32: em\_hcg\_other: \*  
33: em\_hcg\_mus: \*  
34: em\_hcg\_pln: \*  
35: em\_hcg\_rtd: \*  
36: em\_hcg\_mam: \*  
37: em\_hcg\_vtc: \*  
38: em\_ay: \*  
39: em\_hlgo\_hum: \*  
40: em\_hlgo\_mus: \*  
41: em\_hlgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1454.2	99.9	1455	8 AY281377	AY281377 Hypocrea
2	481.8	33.1	1900	8 AB003109	AB003109 Humicola
3	452.2	31.1	1695	8 AY081764	AY081764 Talaromyc
4	397.6	27.3	2171	8 AB003110	AB003110 Hypocrea
5	373.4	25.7	1590	8 AY434988	AY434988 Trichoder
6	278.2	19.1	1828	8 AK119546	AK119546 Oryza sat
7	278.2	19.1	2032	8 AY129294	AY129294 Oryza sat
8	253.2	17.4	1988	8 AF268911	AF268911 Aspergill
9	248.6	17.1	2045	8 AK067001	AK067001 Oryza sat
10	247.8	17.0	2068	8 AF439322	AF439322 Talaromyc
11	246	16.9	1404	6 AX654334	AX654334 Sequence
12	221.6	15.2	2148	8 AK067231	AK067231 Oryza sat
13	221.6	15.2	2148	8 AK120430	AK120430 Oryza sat
14	215.8	14.8	1811	8 AK066710	AK066710 Oryza sat
15	214.6	14.7	1503	6 AX652977	AX652977 Sequence
16	213	14.6	1797	8 AY056828	AY056828 Oryza sat
17	212.2	14.6	302007	1 SC0939132	AL939132 Streptomy
18	210.4	14.5	1689	8 CSAE26G	D83177 Costus spec
19	209	14.4	1665	8 AK104707	AK104707 Oryza sat
20	205.2	14.1	1792	8 AK105908	AK105908 Oryza sat
21	203.6	14.0	1792	8 AK089938	AK089938 Oryza sat
22	202	13.9	1803	8 AK059210	AK059210 Oryza sat
23	201.8	13.9	1857	8 AK103027	AK103027 Oryza sat
24	200.2	13.8	1917	8 AK105026	AK105026 Oryza sat
25	198.6	13.6	1894	8 AK100165	AK100165 Oryza sat
26	198.6	13.6	1909	8 OSU28047	U28047 Oryza sat
27	196.4	13.5	1842	8 AK071372	AK071372 Oryza sat
28	194.8	13.4	1824	1 CFGLUCBET	M94865 Cellulomona
29	191	13.1	2081	1 SSBGLJGB	229625 Streptomy
30	188.6	13.0	298450	1 SC0939107	AL939107 Streptomy
31	186.2	12.8	5996	1 AF086819	AF086819 Thermobif
32	183.2	12.6	1796	8 AK071058	AK071058 Oryza sat
33	182.8	12.6	1734	6 AX412252	AX412252 Sequence
34	182.8	12.6	1734	6 AX506587	AX506587 Sequence
35	182.8	12.6	1734	6 AX651286	AX651286 Sequence
36	182.8	12.6	1881	8 AF159376	AF159376 Arabidops
37	180	12.4	6219	4 OCLPH	X07995 Rabbit mRN
38	178.6	12.3	1533	6 AX653834	AX653834 Sequence
39	178.6	12.3	6183	4 OCLPH2	227167 O. cuniculus
40	178.2	12.2	1562	8 AY142610	AY142610 Arabidops
41	178.2	12.2	2016	6 AX454762	AX454762 Sequence
42	178.2	12.2	2016	6 AX464120	AX464120 Sequence
43	178.2	12.2	2016	6 AX491240	AX491240 Sequence
44	178.2	12.2	2016	9 AY358729	AY358729 Homo sapi
45	178.2	12.2	2266	6 AX574385	AX574385 Sequence

## ALIGNMENTS

RESULT 1  
LOCUS AY281377 1455 bp mRNA linear PLN 15-JUN-2003  
DEFINITION Hypocrea jecorina strain Qm6a Cel1b (cel1b) mRNA, complete cds.  
ACCESSION AY281377  
VERSION AY281377.1 GI:31747169  
KEYWORDS  
SOURCE  
ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Hypocreales; Hypocrea.  
REFERENCE  
1 (bases 1 to 1455)  
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Coleman, N.S., Goedegebuur, F., Houtek, T.D., Engstrand, G.J.,  
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,

Pred. No. is the number of results predicted by chance to have a

Oliveres, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.  
Transcriptional Regulation of Biomass-Degrading Enzymes in the  
Filamentous Fungus *Trichoderma reesei*  
J. Biol. Chem. (2003) in press  
2 (bases 1 to 1455)  
Foreman, P.K., Brown, D., Dankmeyer, L., Dean, R., Diener, S.,  
Dunn-Colman, N.S., Goedegebuur, F., Houlek, T.D., Mitchell, T.,  
Oliveres, H.A., Teunissen, P.J.M., Yao, J. and Ward, M.  
Direct Submission  
Submitted (23-APR-2003) Genencor Intl., 925 Page Mill Road, Palo  
Alto, CA 94304, USA  
Location/Qualifiers  
1..1455  
/organism="Hypocrea jecorina"  
/mol\_type="mRNA"  
/strain="QM6a"  
/db\_xref="taxon:51453"  
/note="anamorph: *Trichoderma reesei*"  
1..1455  
/gene="cellb"  
1..1455  
/gene="cellb"  
/note="putative beta-glucosidase"  
/codon\_start=1  
/product="Cellb"  
/protein\_id="AAP5758.1"  
/db\_xref="gi:31747170"  
/translation="MPSIALPNDPEMFAPATAVOIRGAVKEGGRPSIMPTCHLP  
SRITGANGDVACDHYRDEPDLITTKGAKAYRFSISMRIILPGLRDLVYNEGLEP  
FYSKLIDALNRGITTPWVTLVHMDLPOLHRYGAMLVESVQJDFEYALCEEREG  
DRVONMTINKPWIQAIYGVATGSNAPERSINRSTEGNTATPEWLAGAQIWSHAR  
AVAYSRDFRPSQGOIGISLNGDYEPEDWSDNEPRDEKAEARMEFIHGMANP1FLK  
KDYESMKKOLGEPLALTPADPAFLINAGTFDFPMNYTOSFARHDLPEVETDYG  
AIHEHKNKDGSPVGESEGLAWLRSCPMPEKHLARVGLYKDPILYTENCCPGRGE  
NMTEBEAVNDPFRIRYRPSHLDISKAITQGVVYKGFAMALLDNLEMSGYPRRG  
VTFDYTYLTKRTPKKSALVLKMPAARQVAVAA"

ORIGIN

Query Match 99.9%; Score 1454.2; DB 8; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 2,66-262;  
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTGGTACTGCTCTGCCCAACGACTTTGAATGGGGCTTGCACGCGCGCTTAC 60  
Db 1 ATGCCGAGTGGTACTGCTCTGCCCAACGACTTTGAATGGGGCTTGCACGCGCGCTTAC 60  
QY 61 CAGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCGCGCTTCCATCTGGGACACCTACTGC 120  
Db 61 CAGATCGAAGCGCGCTCAAGAAAGTGGCGCGCGCGCGCTTCCATCTGGGACACCTACTGC 120  
QY 121 CACCTGAGACCATGCGGACCAACGCGCGCAACGCGGATGTGGCTTGGATCACTACCAAC 180  
Db 121 CACCTGAGACCATGCGGACCAACGCGCGCAACGCGGATGTGGCTTGGATCACTACCAAC 180  
QY 181 CGCTACGATGAGAGACTTTGATCTTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCC 240  
Db 181 CGCTACGATGAGAGACTTTGATCTTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCC 240  
QY 241 TTGTGCTGTGGCGGATCACTTCCCTCGGCGGCAAGCTGAGATCCCGTCAACGAGAGGGA 300  
Db 241 TTGTGCTGTGGCGGATCACTTCCCTCGGCGGCAAGCTGAGATCCCGTCAACGAGAGGGA 300  
QY 301 ATTGAGTTTTCAGAGAACTGATGACGCGCTGTGAGAGCGGGGATACACGCTTGGGTG 360  
Db 301 ATTGAGTTTTCAGAGAACTGATGACGCGCTGTGAGAGCGGGGATACACGCTTGGGTG 360  
QY 361 ACTTTGACCACTGGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGGCTGAGCTCAAC 420  
Db 361 ACTTTGACCACTGGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGGCTGAGCTCAAC 420  
QY 421 GTGGAAGAGTCTCACTGGAATTTGAGCGGCTATGAGAGGTTGCTTTGAAAGTTTGGG 480  
Db 421 GTGGAAGAGTCTCACTGGAATTTGAGCGGCTATGAGAGGTTGCTTTGAAAGTTTGGG 480

QY 481 GACCGATCCAGAACTGGATCACCATCAACGACCCTGGATTGAGGCGCATCTATGATAT 540  
Db 481 GACCGATCCAGAACTGGATCACCATCAACGACCCTGGATTGAGGCGCATCTATGATAT 540  
QY 541 GCCACCGGCGACGACCGCCCGGCGGACGAGACGACTTAAACAAGACTCCACCGAGGGCAAC 600  
Db 541 GCCACCGGCGACGACCGCCCGGCGGACGAGACGACTTAAACAAGACTCCACCGAGGGCAAC 600  
QY 601 ACGGCCACTGAGCGGTGGCTCGCTGGAAGAGCCCAAGTCAATGAGCCATGCGCGCGCGT 660  
Db 601 ACGGCCACTGAGCGGTGGCTCGCTGGAAGAGCCCAAGTCAATGAGCCATGCGCGCGCGT 660  
QY 661 GCCGTTCACAGCAGGGAATTTGCGCCCTGCAAAAGGGCCAGATCGGCATCTGCTCAAC 720  
Db 661 GCCGTTCACAGCAGGGAATTTGCGCCCTGCAAAAGGGCCAGATCGGCATCTGCTCAAC 720  
QY 721 GGGCACTACTATGAGCCCTTGGGACAGCAATGAGCCTTGGGACAAAGAGGCTGCTGAGCGA 780  
Db 721 GGGCACTACTATGAGCCCTTGGGACAGCAATGAGCCTTGGGACAAAGAGGCTGCTGAGCGA 780  
QY 781 CGGATGGAATTTTCAATTTGGCTGTTGCCAATCCCATCTTCTTGAAGAAGACTATCCA 840  
Db 781 CGGATGGAATTTTCAATTTGGCTGTTGCCAATCCCATCTTCTTGAAGAAGACTATCCA 840  
QY 841 GAGAGCATGAAGAGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCC 900  
Db 841 GAGAGCATGAAGAGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCC 900  
QY 901 ATCTCTCAATGCGGAGAGACCGACTTCTACAGCGATGAATTAATCAATCCAGTTGGCG 960  
Db 901 ATCTCTCAATGCGGAGAGACCGACTTCTACAGCGATGAATTAATCAATCCAGTTGGCG 960  
QY 961 CGCCACCTAGACGGTCCCGTCCCGGACGAGCACTATCTCGGCGGCATCTCAGACCGACAG 1020  
Db 961 CGCCACCTAGACGGTCCCGTCCCGGACGAGCACTATCTCGGCGGCATCTCAGACCGACAG 1020  
QY 1021 GAGAAATGAAGACGCGACGCCCGTTGGCGAGAGAGCGGCTTCCGCTGCTGCTCTGCTC 1080  
Db 1021 GAGAAATGAAGACGCGACGCCCGTTGGCGAGAGAGCGGCTTCCGCTGCTGCTCTGCTC 1080  
QY 1081 CCGGACATGTTCCGGAAGCATCTTGGCCGGGTGTAACGGCTGTACAGGCGCATCTAC 1140  
Db 1081 CCGGACATGTTCCGGAAGCATCTTGGCCGGGTGTAACGGCTGTACAGGCGCATCTAC 1140  
QY 1141 ATCACCGAAGACGATCCCGTGCCTGAGAGAGGAGAACATGACGTGAGAGAGGCGCTC 1200  
Db 1141 ATCACCGAAGACGATCCCGTGCCTGAGAGAGGAGAACATGACGTGAGAGAGGCGCTC 1200  
QY 1201 AAGCACTCTTCCGATCCGATCTTTGACTCGCACTTGAATCTGATTTTCCAAGGCCATT 1260  
Db 1201 AAGCACTCTTCCGATCCGATCTTTGACTCGCACTTGAATCTGATTTTCCAAGGCCATT 1260  
QY 1261 ACCCAGACGCGCGTCTGCTCAAGGGGTACTTTGCGGGGGGTGCTGATTAACCTGGAA 1320  
Db 1261 ACCCAGACGCGCGTCTGCTCAAGGGGTACTTTGCGGGGGGTGCTGATTAACCTGGAA 1320  
QY 1321 TGCTCAATGAGCTACGAGACCGAGATTGCGCGCTCACTGTTCAACAGACTACACCACTCAAG 1380  
Db 1321 TGCTCAATGAGCTACGAGACCGAGATTGCGCGCTCACTGTTCAACAGACTACACCACTCAAG 1380  
QY 1381 CGCAGCGCCCAAGAGTCTGCTGCTCAAGAGCAATGTTGCGGCGCGCGCAGAGGGTT 1440  
Db 1381 CGCAGCGCCCAAGAGTCTGCTGCTCAAGAGCAATGTTGCGGCGCGCGCAGAGGGTT 1440  
QY 1441 AAAAGTGGCGGATTA 1455  
Db 1441 AAAAGTGGCGGATTA 1455

RESULT 2  
AB003109  
LOCUS  
AB003109 1900 bp DNA linear PLN 25-DEC-2002

DEFINITION	Humicola grisea var. thermoidea bgl4 gene for beta-glucosidase, complete cds.
ACCESSION	AB003109
VERSION	AB003109.1 GI:4249559
KEYWORDS	beta-glucosidase.
SOURCE	Humicola grisea var. thermoidea
ORGANISM	Humicola grisea var. thermoidea
REFERENCE	Eukaryota; Fungi; Ascomycota; microsporitic Ascomycota; Humicola.
AUTHORS	1 Takashima, S., Nakamura, A., Hidaka, M., Masaki, H. and Uozumi, T.
TITLE	Molecular cloning and expression of the novel fungal
JOURNAL	beta-glucosidase genes from Humicola grisea and Trichoderma reesei
AUTHORS	J. Biochem. (1999) In press
TITLE	2 (bases 1 to 1900)
FEATURES	Uozumi, T.
Source	Direct Submission Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo, Department of Biotechnology, Faculty of Agriculture; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan (E-mail: uozumi@mc.b.t.u-tokyo.ac.jp, Tel:03-5684-0387, Fax:03-5684-0387)
gene	Location/Qualifiers
CDS	1..1900
	/organism="Humicola grisea var. thermoidea"
	/mol_type="genomic DNA"
	/sub_species="IPO9854"
	/db_xref="taxon:5528"
	join(289..332,437..1823)
	/gene="bgl4"
	join(289..332,437..1823)
	/gene="bgl4"
	/codon_start=1
	/product="beta-glucosidase"
	/protein_id="BAJ74998.1"
	/db_xref="GI:4249560"
	/translation="MSLPDPFKMGFPATPAVOIEGSVNEDGRPSIMTDFCALPEKIID GSSGAVCDYSYKRTKEDIALKELGANSYRSISWSRIPLGSGNDINOGKIDHYK FVDDILFAGITPPILPTLFHMDLPDALDRYGGFLKEEFADFENYARIMPKALPKCHG WITNEPMCSAILGNTGYFAPGHTSRSKSPVDSAREPVIYHNILIHARAVKAY REDKPTGGEIGITVALLDPAATLPMDPDPADIECDKIEPAISWPAIDPIYFGKYPDS MRKQDRLPREFTEBYVALYKGSNDFPGAHNTYANYIKHKTGVPEDDPLGNLETLFY NKYDCTGPEPTOSFWLRPHAGCFRDLNMLSKRYGPRIYVTENGTSLKGNDFPLEQ VLEDFPVKXFNDYVRAAAMAAVAEDGDNVIRGYLAMSLLDNFEMAGYETRFYGVYDY ANDKRPVKSASAKLPELPSLIRKE"
ORIGIN	
Query Match	33.1%; Score 481.8; DB 8; Length 1900;
Best Local Similarity	61.7%; Id. Mismatches 513; Indels 15; Gaps 5;
Matches	854; Conservative 1;
QY	47 CAACGGCGCGCTTACAGATCGAAGCGCGCTCAAGAAAGTGCGCGCGCCGCTCATCT 106
DB	427 CAATCTCAGATTCACGATTCAGAGGCTTCCTCAACGAGATGCGCTGCGCTCATCT 486
QY	107 GGACACGACTGCGCACCTGGAGCCATCGCGACCAACGCGGCCAACGCGCATGTGCTT 166
DB	487 GGACACCTTCTGGGCCATCCCGGCAAGATCGCGACGAGCTCGGCGCGCTGGGCT 546
QY	167 GCGATCTACTACACCGCTACAGATGAGGACTTGAATCTTTTACCAAGTACGCGCCCAAG 226
DB	547 GCGACTGCTACAAACGACCAAGAGGACATTTGCCCTTCTCAAGAGCTCGGCGCAACT 606
QY	227 CCTACCGCTTCTCTTGTGCGTGTGCGCGATCATTTCCCTCGGCGGCGAGCTGATCCCG 286
DB	607 CGTACCGCTTCTCATCTCGTGTGTGCGCATCATCCCTCGGGGGCGCGCAATACCCCA 666
QY	287 TCAACGAGGAGGAATTGATTTTACAGCAAACTGATTTAGCGCCCTGTGAGCGGGGTA 346
DB	667 TCAACGAGGAGGAGATCGACCACTACGTCAGTTGTTCAGACGACTGTCAGGCGCGCA 726
QY	347 TCAGCGCTTGGGTGCTTTGTACACATGGGATCTGGCTCAGGCGCTTACAGATGGCTATG 406
DB	727 TCAGCGCTTATCATCCTCTTTCACAGTGGAGCTGGCGGAGCCCTCGACCAAGCGTTACG 786

QY	407	GAGGCTGGCTCAAGSTGAAAGAGGTCCAGCTGGACTTTGAGCGGTAATGCGAGGTTGCT	466
Db	787	GGGGCTTCTTCACAGAGAGAGATTGCCCGCCCACTTTGAGAACTACGCCCCGATCATGT	846
QY	467	TTGAACGTTTTGGGAGCCGAGTCCAGACTGATATCAACATCAAGAWCCCTGATTCAGG	526
Db	847	TCAAGGCCATCCCAAGTG--CAAGCATGATATCACTTCAAGAGCCCTGGGCTCCG	903
QY	527	CCATCTATGATATAGCCACCGGCGAGCAAGCCCCCGGCGAGAGACAGATTTACAGACT	586
Db	904	CCATCTTCGGGTACACAGCGGCTATTTCCGCCCCCGGCCACACTTCGAGCCGACGCAAGT	963
QY	587	CCACCGAGGGCAACATGCGCACTGAGCGCTGAGCTCGGTGAAAAGCCCATATGAGCC	646
Db	964	CCCCCGTCGGCGAGAGCGCCCGCGAGCCCTGATTTGTCGGCCACACATCTCTATCGCCC	1023
QY	647	ATGCCCGCGCGTGGCCGCTTACAGCAGGAGACTTTCGCCCCCTCGCAAAAGGGCCAGATCG	706
Db	1024	ACGCCCGCGCGTCAAGGGCTTACCGCGAGGACTTCAAGCCACCCACAGGCGGGAGATCG	1083
QY	707	GCATCTGGCTCAAGGGGACTACTATAGCCCGTGGGACACCAATAGCCCTCGGACAGG	766
Db	1084	GCAATCACCTTGAAAGGAGCGCACCCCTGCGTGGAGCCCGAGAGACCCGAGCAGACTTG	1143
QY	767	AGGCTGTGACGACGCGATGAAATTTTCACTTGGCTGTGTCATATCCCATCTTCTTGA	826
Db	1144	AGGCTGTGACGACGCGAATGGAATGCCCATCTCTGTGTTTGGCGACACATCTCTATCGCTG	1203
QY	827	AGAGAGCTATCCAGAGAGATGAAAGAGAGCTGGCGAGAGGCTTTCAGCCTCACTC	886
Db	1204	GCAAGTAC--CCGACTCATGTGGCAAGAGAGCTGGCGAGCCCGCTCCGAGTTCACTC	1260
QY	887	CCGGGACTTTGGCATCTCTCAATGCCGAGAGACCGACTTCTACGGGATGAATTAATCA	946
Db	1261	CCGAAAGGTGCGCCTTGCTCA--GGGCTTCACGACTTCTACGGGATGAACCATCA	1317
QY	947	CATCCCAATTGCGCGCGCACTAGACGATCCCGTCCCGGAGAGGCACTATCTCGGCCCA	1006
Db	1318	CGGCACTATCATACAGCACAAAGCGGGCGTGGCGCCCGAGAGACGACTTCTCGGGCAAC	1377
QY	1007	TCCATGAGACACAGAGATTAAGACGGCGAGCCCGTTGGCGAGAGAGCGGCTCGCT	1066
Db	1378	TCCAGAGGCTCTTTTACAAAGTAACGGCACTGCACTGGCCCGGAGAACCACTCTTCT	1437
QY	1067	GGCGGCTCTCGCCGGGACATGTTCCGGAAGCATCGGCCGGGTATAGGCTCTAGC	1126
Db	1438	GGCTGCGCGCGACGCGCCAGGGCTTCCGCGACTGCTCAACTGGCTCAGCAGCGCTAGC	1497
QY	1127	GCAAGCCC--ATCTACATCAACGAGACGAGATGCCCCGTCCTCGAGAGAGAAATGA	1183
Db	1498	GCTACCCCAAGATCTTAGTGACGGAACCGAACCTTCGCTCAAGGGGAGAAACGACTGC	1557
QY	1184	CGTCCGAGAGAGCGGCTCAAGACCCCTTCCGCACTCGSTACTTTGATCTCGCATCTTGA	1243
Db	1558	CCCTCGACAGGTCTCTGAGAGACCATTCGCGCTCAAGCTTCTCAAGACTCACTCGGCG	1617
QY	1244	CGATTTTCAAGGGCATTAAACCGAGAGCGGCGTGGTGGTCAAGGGGTACTTTTGGCGGGCGT	1303
Db	1618	CCATGGGCGCGCGCTGCGCGAGAGACGGCTGCAAGTTCGCGGTTACTCGGCTGGTCCG	1677
QY	1304	TGCTCGATTAATTGGAATGCTCAGATGGCTACGGAACCAATTCGCGGCTCAAG	1363
Db	1678	TGCTCGAACACTTTGAGTGGGCGGAGGCTACGAGACGAGAGTTTGGGCTGACCTATGTGG	1737
QY	1364	ACTA---CACCACTTCAAGCGCACGCCCAAGAAAGTTGCGCTGTCTCAAGACATGT	1420
Db	1738	ACTATGCCAACGACCAAGAGGTATCCAAAGAGTGGCCCAAGAGCTCAAGCCGTTGT	1797
QY	1421	TTG 1423	
Db	1798	TTG 1800	

RESULT 3  
AY081764 1695 bp mRNA linear PLN 31-MAY-2002  
LOCUS Talaromyces emersonii beta-glucosidase (bg11) mRNA, complete cds.  
DEFINITION  
ACCESSION AY081764  
VERSION AY081764.2 GI:21284364  
KEYWORDS  
SOURCE Talaromyces emersonii  
ORGANISM Talaromyces emersonii  
REFERENCE 1  
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
TITLE The molecular cloning of the beta-glucosidase genes of Talaromyces emersonii  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1695)  
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
TITLE Direct Submission  
JOURNAL Submitted (04-MAR-2002) Biochemistry, National University of Ireland, Galway, Galway, Ireland  
REFERENCE 3 (bases 1 to 1695)  
AUTHORS Murray, P.G., Collins, C.M. and Tuohy, M.G.  
TITLE Direct Submission  
JOURNAL Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Galway, Ireland  
REMARK Sequence update by submitter  
COMMENT On May 31, 2002 this sequence version replaced gi:19568099.  
FEATURES  
source location/Qualifiers  
1..1695  
/organism="Talaromyces emersonii"  
/mol\_type="mRNA"  
/db\_xref="taxon:68825"  
1..1695  
/gene="bg11"  
/locus\_tag="BG\_11"  
/note="family 1 glycosyl hydrolase"  
/codon\_start=1  
/product="beta-glucosidase"  
/protein\_id="AA09551.2"  
/db\_xref="GI:21284365"  
/translation="MTTETPTTTTSTLPPDLWGFATASYOIEGAFDEDEGLPSIM  
DTEKTPGVKVDGNDVACDSYHRTBEDILKKGAKYRSLSPRIIPGGRND  
PIKEKQDFYKPLDLDLAAGIEPFTLFDLDELMSKYGKLNKEEVAAYANA  
RVVFNALGSKVKHMTENBPMGSGVLGHTNGKAPGRTSPRTSGPDEGIREPWIYH  
NLVAAGTVVDIYREFEKOGEGEIGTLTGDAEAPWDPENPDVRECDPKIEFALSM  
PADPIYHGKYPDSVWQDGLPKPTPEBEIAPFHGSDPYGMNHYCENTYRNRTGEED  
PEDIAGNLDLMEBKNGNPFGPTQCEMLRPFPLGFKLKLMLADRNNPKIYVTEEG  
TSVAGSDKPLEEVLNDEFVQYRDYIGAMVDAVADGVNVAAYAMSLDNFEWSE  
GYRSEFVTYVDYKNGQKRIPKSAIVIGELFNKYIRKE"  
ORIGIN  
Query Match 31.1%; Score 452.2; DB 8; Length 1695;  
Best Local Similarity 59.5%; Pred. No. 1.1e-74;  
Matches 833; Conservative 1; Mismatches 559; Indels 12; Gaps 4;  
QY 17 CTCTGCCCAAGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATGGAAGCGCGG 76  
Db 147 CTCTTCTCCGACTTCTCTGGGGTTTCGCACTGCACTACGATGGAAGGGCTT 206  
QY 77 TCMAAGAAGTGGCGCGCCCGCTTCATCTGGGACACGTAATGCGACCTGGAAGCATCGC 136  
Db 207 TCAGACGAGATGGCGCTTCTCGATATGGGACATTTTCAGCAAGACGCCGCGCAAGG 266  
QY 137 GCACCAACGCGCCCAAGCGGATGGCTTGCGATCACTACCCGCTAGATGAGAGACT 196  
Db 267 TCAGACGCGGACCAATGCGGATGTCCCTGCGATCTGTAACACCGATCCGATGAGACCA 326  
QY 197 TTGATCTCTTGAACCAATGACGCGCAAGGCTTACCGCTTCTCTGTCGTGGTCGCGA 256  
Db 327 TTGAGATATTGAAGAATATACGTCGCAAGATCTATCGCTTCTGTTGCTGCGCAAGTA 386

QY 257 TCATCCCTTCGCGCGACGGCTGATCCGTCACAGAGAGGAATTGATTTTACACA 316  
Db 387 TCATCCCTTCGCGCGCGACGGCTGATCCGTCACAGAGAGGAATTGATTTTACACA 446  
QY 317 AACTGATTTGAGCGCCCTTGAAGGCGGATTCACGCTTGGGTGACTTTGTACCACTGGG 376  
Db 447 AGTTCTTGATGATTTACACGCGCGGCGATCAGAGCCCTTGTCAGCTTCCACTGGG 506  
QY 377 ATTCGCTCAGCGCTTCAAGATCGCTATGAGAGCTGCTCAACGTGGAAGAGTTCAGC 436  
Db 507 ATTCGCGCGACGAGCTGATGAAAGCGATGAGGGGACAGCTCAACAGAGAGAGTTGTGG 566  
QY 437 TGAATCTTGAAGCGGTATGACAGGTTGTGCTTTGAACCTTTTGGGACCGAGTCAGACT 496  
Db 567 CCGACTAGCGCAACTATGACCGCGGTGTCTTCAACGCGCTCGAGAGGTCAAGACT 626  
QY 497 GATACACATCAACGAMCCCTGATTCAGGCACTATGATATGCAACCGGACGAGCAAG 556  
Db 627 GATACATTTCAACGAGCCCTGATGATTCGCTGCTGCGCCCAACACCGGACGAGCA 686  
QY 557 CCGCGGCGAGAGCAGCATTTACAGCACTCCAGAGGCAACACTGCACTGAGCCGT 616  
Db 687 CCGCGGCGGAGAGAGCATCGAGCAAAAGTCCGAGGAGAGCGGCACTCGGAACTT 746  
QY 617 GGCCTCGTGAAGAGCGCCAGATCATAGAGCATGCGCGCGCTGCGCTTACAGAGG 676  
Db 747 GATTCGAGGTCAAACTTCTGTGCTCATGATGATCGTTGTGACATCTATCGAGAG 806  
QY 677 ACTTGCCTCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGCACTATAGC 736  
Db 807 AGTTCAAGAGAAACAAGGGGAGAAATGAGATATCACTCAACGCGCATGGCCGAAC 866  
QY 737 CTTGGGACAGCAATGAGCTTCGGAACAAGAGGCTCTGAGCGACGATGAAATTTACA 796  
Db 867 CATGGATCCCGAAACCCAGACAGATCGAGGCTCGCAACCGAAATCGAATTCGCA 926  
QY 797 TTGGCTGTTTGGCAATCCATCTTTTGAAGAGCACTATCCAGAGACATGANAAGC 856  
Db 927 TCTCGTGTGTTGCGGAGCCGATTTACACGGGAAATAC--CCGACACATGTTTAAAC 983  
QY 857 AGCTGGCGAGAGAGCTTCAGCCCTCATCTCCGCGACCTTGGCATCTCAATGCGGAG 916  
Db 984 AGTTGGCGACCGTCTGCGGAAGTTACCGCGGAAGAAATCGCTTCGGA--CGCA 1040  
QY 917 AGACGCACTTCAAGCGCAATTAATTAACATCCAGTTCCGCGCCACTAGACGGTC 976  
Db 1041 GCAACGATTTTACGCGCATGAACCACTACTGCGAAATCAATCCGCAACCGACCGGCG 1100  
QY 977 CCGTCCCGAGACGACTATCTCGGCGCATCATGACACAGAGAAATTAAGACGCA 1036  
Db 1101 AACCAAGCCCGAGAGCAATGTGCGGACCTGAGATCTGTATGAGAGCAAGAAACGCA 1160  
QY 1037 GCGCCGTTGGCGAGAGAGCGGCTCGGCTGCGCTGCTCTCCCGGACATGTTCCGA 1096  
Db 1161 ACCCATTCGCGCCCGACCAATGAGAGTGGCTGCGCGCTTCCCGCTGGGGTCAAGA 1220  
QY 1097 AGCATCTCGCCCGGGGTATCGGCTGTACGGAAGCC--ATCTACATCAACGGAAGC 1153  
Db 1221 AGCTGCTGAAGTGGCTTGGAGACCGGTATCAACACACCCCAAGATCTACGACCGAAGC 1280  
QY 1154 GATGCCCGTGCCTGAGAGAGAAATGACCTGCGAGAGGCGCTCAACGACCCCTTCC 1213  
Db 1281 GAGACGAGCGTCAAGGCGGAGAGAGCAAGCGCTGAGAGAGTGTCAACAGATGAGTTCC 1340  
QY 1214 GCATCGGATCTTTGATCTGCACTTGAAGCTGATTTCCAAAGGCAATTAACCAAGAGCGCG 1273  
Db 1341 GGTGCAAGTAACTACCGGAGCTATATCGGCGGAGATGATGCTGTTCCACAGATGAGG 1400  
QY 1274 TCGTGTCAAGGAGTAACTTTGGTGGCGCTTGCATGATTAATGGAATGATCAAGTGGCT 1333  
Db 1401 TCAAGCTCAAGGCTTACATGCTTGGAGTTGTTAGACAACTTGGAAATGATCGAAGGCT 1460



Oy		1334	ACGCAACCAGATTGGCGGCACGTTCACAGACTAACACACC---	CTCAGAAGCACAGCCCA	1390
Dd		1461	ACCGTCCAGATTCGGAGTAGCTCTGCTGCATACAAAGACGGCAGAAAGAGATTCCGA		1520
Oy		1391	AGAACTGTGCCCTGTGCTCTCAAGCATGTT	1421	
Dd		1521	AGAAAGTGCTCTTGTCTCATCGGGAGCTGTT	1551	
RESULT 4					
AB003110		2171 bp	DNA	linear	PLN 25-DEC-2002
LOCUS	Hypocrea jecorina bg12 gene for bete-glucosidase, complete cds.				
DEFINITION	AB003110				
ACCESSION	AB003110.1	GI:4249561			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	Hypocrea jecorina Hypocrea jecorina Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.				
REFERENCE					
AUTHORS	Takahima,S., Nakamura,A., Hidaka,M., Masaki,H. and Uozumi,T.				
TITLE	Molecular cloning and expression of the novel fungal beta-glucosidase genes from Humicola grisea and Trichoderma reesei				
JOURNAL	J. Biochem. (1999) In press				
REFERENCE	Uozumi,T. 2 (bases 1 to 2171)				
AUTHORS	Direct Submission				
TITLE	Submitted (09-MAR-1997) Takeshi Uozumi, The University of Tokyo, Department of Biotechnology, Faculty of Agriculture; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan [E-mail : uozumimcb.bt.a.u-tokyo.ac.jp,				
JOURNAL	Tel:03-5684-0387, Fax:03-5684-0387)				
FEATURES					
source	Location/Qualifiers 1..2171 /organism="Hypocrea jecorina" /mol_type="genomic DNA" /strain="QM9414" /db_xref="taxon:51453" /note="anamorph: Trichoderma reesei" join(556..596,670..2029) /gene="bg12" join(556..596,670..2029) /gene="bg12" /codon_start=1 /product="beta-glucosidase" /protein_id="BAA74959.1" /db_xref="GI:4249562"				
gene	/translation="MLPRDPMWGFRATAAYVLEGAVDDQGRPSITWDFCAOPKTIADG SSGYACDSYKRIALDELALSLSGAKSTRESISRIIPBGEGDAVNQAIDHYVR VTDLADAGITPFTLPHWDLEGLHQRGVLNTEFLDENYARVMFRLPVRW IDTFEPULSAIPIGVSGTFAPRGROSTSPMTVGHNIIIVAHGRAVKAADPKPASGDG OIGVLNDFTYPMWDADPADKEAEERLEFFTWAFADPIYLGDYPASMRKQLDRLE TFPEERLVHGNSDNFYGMNHVTSNYLRHSPPASADDTGNVDVLFNNKGNCIGPE TGSWMRRCAGRPDPFLVMISKRYGPITYTENGTSKESDLPKKEILEDDEPRVKK YNETRIMAVTAVELDGVNVKKYFMASLMDNEMADGVTRGVTYVDVENQKRPFKLA SAKLKPLFDLIAAA"				
CDS					
ORIGIN					
Query Match	27.3%; Score 397.6; DB 8; Length 2171;				
Best Local Similarity	59.4%; Pred. No.1.7e-64;				
Matches	824; Conservative 1; Mismatches 515; Indels 48; Gaps 77				
Oy		45	CGCAAACGGCCCCCTTCACGATTCGAAGGGCGGTCAAAAGAGTGACCGCGCCGCTCAT	104	
Dd		658	CCCCTCCGACACGCTACGATCGAGTGAAGGGCGCCGTGCAACAAGACGGCCGCCAGAT	717	
Oy		105	CTGGGACACGTACTGCACTTGAGCCATCGCGCACCAACGCGCCAAAGCGCATGTGGC	164	
Dd		718	CTGGGACACGTTCTTGCGCGGACCGCCGGAAGATCGCCGACGCGCTGTGCGGCGTGA	777	
Oy		165	TTGGCATACATCAACCGCGCTACGATGAGAGACTTTGATCTCTTGACCAAGTACGGCGAA	224	
Dd		778	GTCGACTCTGTACAACCGCAGCGCCGAGACATTTGCGCTGCTCAAGTCTCTCGGGGCCA	837	

OY	225	GGCCATCCGGCTTCTCTGTGTGATGCTGCGGATATATTTCCCTCGGCGGACAGGCTTGATATCC	284
Db	838	GAGCTACCGGCTTCTTCATCTGTGTGTGCGCATATATCCCGAGGGCGGGCGCGGCATATCC	897
OY	285	CGTCAACGAGGAGGAAATTGATGTTTACAGCAACGTATGACGCGCTGTTGAGGCGGGG	344
Db	898	CGTCAACGAGGCGGCGCATATCGACCACTAGTCATGTTGTTGTCGACGACTGCTCGACGCGG	957
OY	345	TATACGCGCTTGGGTGACTTTGTATCACTGGGATCTGTCTCAGGCGCTTACAGATTCGCTA	404
Db	958	CATACACGCGCTTTCATCACCCCTTTCACATGGGACCTCTCCGAGGCGCTGATCGAGCGTA	1017
OY	405	TGAGAGCGTGGCTCAACGCTGGAAGAGATGTCACAGCTGGAATTTGAGGCGGTATGCGAGTGTG	464
Db	1018	CGGGGGGCTGTGTAACCGCACCGAGTTCCGCTGACTTTGAAAACTACGCGCGCT---	1074
OY	465	CTTTGAACGTTTTGGGGAACCGAGTCCAGAACTGATGATCAACATCAAGACCCTGATTTCA	524
Db	1075	CATGTTCAAGGGCGCTGCGCCAAAGTGCGCAATGATGATCACTTCAACGAGCGCGTGTCTC	1134
OY	525	GGCCATCTATGGAATATGCAACCGGCGACAAAGCCCCGGGCGAGGACAGCATTAACAAACA	584
Db	1135	GGCCATCCGGGCTACGCTCTCGGACCTTGCAGCCCGGCGGAGAGCA-----	1184
OY	585	CTCCACGAGGGCAACACTGCACTGAGACGCTGCTCGTTGAAAAGGCCGATCATGAG	644
Db	1185	-----CCTGGAGACGTGAGACGCTGCGCCACAAACATCTCTGTGCG	1224
OY	645	CCATGCCCGCGCGGTGGCCGTCTACAGCAGGGACTTTTGCGCCCTCGCAAA--AGGCCCA	701
Db	1225	CCAGCGCGCGCGCTCAAGGGCGTACCGGACGATCTTCAAGGCCCGCCAGCGGCGACGGCCA	1284
OY	702	GATGGGATCTTCGCTCAACCGGCGACTATGAGACCCCTGGGACAGCAATGAGCCTCGGGA	761
Db	1285	GATGGGACTGTCTCTCAACGGCGACTTACCTACCCCTGGGACGCGCGGACCCGCGCA	1344
OY	762	CAAGGAGGCTCTGAGCGACGGAATGGAATTTCACTTGGCTGTGTTGCCAATCCCATCTT	821
Db	1345	CAAGGAGGCGGCGGAGCGGCGCTCGAGTTCTTACGCGGCTGTGTTGCGGATTCATCTA	1404
OY	822	CTTGAGAGAGACTATCTCAGAGACATGAAAGACAGCTGGGCGAGAGCTTCCAGCCCT	881
Db	1405	CTTGGGC---GACTACCGGCGGTGATGCGCAACACTGGGCGACGCGCTGCGGACCTT	1461
OY	882	CATCCCGCGGACTTTGGCATCTTCATATGCCGAGAGACCGSACTTTTACGSCATGAATTA	941
Db	1462	TACGCGCGAGGAGCGCGCTCTGTGCTCA--CGGCTCCAAAGACTTTTACGSCATGAATTA	1518
OY	942	CTACACATCCCGAGTTGCGCGCGACACTAGAGCGATCCGCTCCGCGAGACGGACTATCTCGG	1001
Db	1519	CTACACGCTCAACTACATCCGCGCACCGGAGTCCGCGCTCCGCGCGACAGCACCGCTGG	1578
OY	1002	CGCCATCCATGAGCACGAGGAAATTAAGACGCGAGCCCGTGGCGAGAGCGGCTT	1061
Db	1579	CAAGTGCACGCTGCTTTCACCAACAAAGGCGCAATGTCATCGGCCCCGCGACGCACTC	1638
OY	1062	GCGCTGGTGGCTCTGCGCCCGGACATGTTCCGAAAGATGTGCGCCGGGGTATCGGCTT	1121
Db	1639	CGCCGTGGCGCGCTGTGCTGCGCGGCTTCCGGACTTCTGTGTGTGATTCAGCAAGAG	1699
OY	1122	GTAAGGCAAA---GCCCATCTACATCACCGAGAACGGAATGCCGTGCCCTGAGAGAGAGAA	1178
Db	1699	GTAAGGCAATCCCGGCATCTACGTAAGAGAGAAAGGACGACGATCAAGGCGAGAGCGA	1758
OY	1179	CATGACGTGCGAGAGGCGCTCAACGACCCCTTCGATCCGATCTTGACTGTGCATT	1238
Db	1759	CTTGCCCAAGAGAAAGATTCTCGAAGATGACTTCAGGAGTCAAGATCACTATAACGATCAT	1818
OY	1239	GAGCTCGATTTCAAGGCAATTAACCGAGACGCGGTGCTGCTCAAGGAGATCTTGTGCTG	1298
Db	1819	CGGTGCAATGATTACCGCGTGAAGCTGAGCGGAGTCAAGCTCAAGGGGATCTTTGCTCTG	1878

Qy 1299 GGCCTTGCTCGATACTTGAGTGTGATGCTACAGCAACCGAATTGGCGTCAAGTT 1358  
Db 1879 GTGCTCATGACACACTTTGAGTGGCGGACGGCTAGTACGAGGTTTGGGGTTACGTA 1938  
Qy 1359 CACAGACT--ACACCACTTCAAGCGCAGCCCAAGAAATCTGCTGCTCAAGGA 1415  
Db 1939 TGTGATTTATGAGATGGGACGAAGCGGTTCCCAAGAAAGACCAAGAGCTTGAAGCC 1998  
Qy 1416 CATGTTTG 1423  
Db 1999 GCTGTTTG 2006

RESULT 5  
AY343988 1590 bp mRNA linear PLN 13-AUG-2003  
LOCUS Trichoderma viride strain AS 3.3711 beta-glucosidase 2 (bg12) mRNA,  
DEFINITION complete cds.  
ACCESSION AY343988  
VERSION AY343988.1 GI:33521681  
KEYWORDS  
SOURCE Trichoderma viride  
ORGANISM Trichoderma viride  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; mitosporic Hypocreales;  
Trichoderma.  
REFERENCE 1 (bases 1 to 1590)  
AUTHORS Liu,B.D., Yang,Q. and Zhou,Q.  
TITLE Cloning and Sequence Analysis of the beta-glucosidase 2 Gene from  
Trichoderma viride AS 3.3711  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1590)  
AUTHORS Liu,B.D., Yang,Q. and Zhou,Q.  
TITLE Direct Submission  
JOURNAL Submitted (15-JUL-2003) Department of Life Science and Engineering,  
Harbin Institute of Technology, Harbin, Heilongjiang 150001, P. R.  
China

FEATURES  
source  
1..1590  
/location=Qualifiers  
/organism="Trichoderma viride"  
/mol\_type="mRNA"  
/strain="AS 3.3711"  
/db\_xref="taxon:5547"  
1..1590  
/gene="bg12"  
192..1544  
/gene="bg12"  
/codon\_start=1  
/product="beta-glucosidase 2"  
/protein\_id="AAQ21384.1"  
/translation="MLPDDPOMGRATAAYOIEGAVDDGRPSIWMFPCAOPKTIADG  
SSGVTACDSYKRTMEDIALKSLGAKSRTFSISRLIIPBGRGDANVQADIDHYK  
VNDLLDAGTIPFIFLPHNDLPEGLHQRGGLNTERPLDENAVRFLPVRW  
ITFNEPLCSAIPGSGGTFAPRQSTSEPTWGHNIIVAHGRAVKAAYRDFKPSGDSG  
QIGVLNGDFTPMDAADPADKEAERLEFFTWAPADPIYLGDPASMRKOLGRLP  
TFPEERLIVHSGNDFGPMNHTSYNIIRHSSPASTPSAPSTOSPSRATASAPR  
SPRLAPRPATSWCGSARGCTPIYLTENSTIKGSBDSIPAKESILDEDPRVKTNYEY  
IRAWTAVELDGVNKGIFAWSLMDNEMADGIVTRGVTVYDENGQKSPRAQRA  
"

ORIGIN  
Query Match 25.7%; Score 373.4; DB 8; Length 1590;  
Best Local Similarity 59.4%; Pred. No. 6.2e-60;  
Matches 802; Conservative 1; Mismatches 497; Indels 51; Gaps 8;

Qy 20 TGCCCAAGCACTTGAATGGGCTTTCGCAAGCGCGCTTACCAATCGAAGCGCCGTCA 79  
Db 196 TGCCCAAGCACTTTCAGTGGGGGTTCCGCAAGCGCGCTTACCAATCGAAGCGCGCGTCA 255  
Qy 80 AAGAAGTGGCGCGCGCGCGCTTCATCTGGACAAGTACTGCCACCTGGAGCCATCGCGCA 139  
Db 256 ACCGAGCGCGCGCGCGCGCGCTTCGGAACAGCTCTGCGCGCAAGCCCGCAAGATCG 315

Qy 140 CCAAGCGCGCAAGCGGATGTGCTTGCATCACTACACCGCTACGATGAGACTTTG 199  
Db 316 CCGAGCGCTGTCGGGGGTGACGCGCGGACCTGCTACACCGCAGCGCGGAGCACTTG 375  
Qy 200 ATCTTTGACCAAGTACGCGCGCAAAAGCGCTTCTCTTGTGCTGCTCGGATCA 259  
Db 376 CGGTGCTCAAGTGGCTGGGGCCAAAGACTACCGCTTCTCATCTCGTGGCGCATCA 435  
Qy 260 TTCCCTCGGGGCGGAGCGTGGATCCGCTCAACGAGGAGGGAATTGATTACAGCAAC 319  
Db 436 TCCCGAGGCGGCGCGGAGTGCCTCAACGAGGCGGAGATGACACTACGTCAGT 495  
Qy 320 TGAATTGAGCGCTTGTGAGCGGGGTATCGAGCTTGGGTGACTTTGTACCATGGGATC 379  
Db 496 TCGTCAACGACCTGCTGACGCGCGGATCAAGCCCTTTCATCACTTTCACGCGGACC 555  
Qy 380 TGCCTCAGGCGCTTACGATCGCTATGAGAGCTGGCTCAACGTGAAAGAGTCCAGCTGG 439  
Db 556 TGCCGAGGCGCTGACATCAGCGGTAAGGGGGCTGCTGAACCGCACCGAGTTCCCGCTCG 615  
Qy 440 ACTTTGAGCGGTATGCAAGGTGTGCTTGAACGTTTGGGGACCGAGTCCAGAACTGA 499  
Db 616 ACTTTGAAAATACGCGCGGCTCATGTTCAGGGCGCT--GCCAAGTGCAGCAACTGA 672  
Qy 500 TCACATCAACAGAWCCCTGATTCAAGGCATCTATGATATGCAACCGGAGCAACGCC 559  
Db 673 TCACCTTCAACGAGCGCGCTGTGCTCGGCGCATCCCGGCTACAGGCTCCGCACTTGGCCC 732  
Qy 560 CGGCGAGGAGCAGCATTTAACAGCACTCCAGAGGCAAACTGCCACTGAGCCGTGGC 619  
Db 733 CCGGCGCGGAGAGCA-----CCTCGAGGCGGTGGA 762  
Qy 620 TCGCTGGAAGGCGCCAGATGATGAGCCAGTCCGCGCGGTGGCGCTACTAGAGGAGCT 679  
Db 763 CCGTGGCGCAACATCTGTGCGCCAGCGCGCGCGCTGCAAGGCGTACCGGACACT 822  
Qy 680 T---TCGCCCTCGCAAAAGGCGCAGATCGGCACTTCTGCTCAACCGGCGCACTATTAGC 736  
Db 823 TCAAGCCCGCAGAGGCGGAGCGGCCAGATCGGCACTGCTCTCAACGGGAGCTTCACTAAC 882  
Qy 737 CTGAGGACAGCAATGAGCTTCTGGGACCAAGAGGCTGTGAGCGACGAGTGAATTACA 796  
Db 883 CTGAGGAGCGCGCGCAGCCGCGGACCAAGAGGCGGCGGCGCTCGAGTCTTCA 942  
Qy 797 TTGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTTCCAGAGAGATGAAGAGC 856  
Db 943 CGGCTGTGCGGATCCATCTACTT--GGGCGACTAACCGGCGTCAATGCGCAAGC 999  
Qy 857 AGCTGGCGAGAGGCTTCCAGCCCTCACTCCCGCGGACTTTGGCATCTCAATGCGGAG 916  
Db 1000 AGCTGGCGGACCGGCTCCGACCTTTAGGCCGAGAGGCGGCGCTCTGCTCA---CGGCT 1056  
Qy 917 AGACCGACTTCTACGCGATGAATTACTACACATCCAGTTCGGCGGCGCACTAGACGCTC 976  
Db 1057 CCAACGACTTTTACGCGATGAACACATACAGTCTCAACTCATCCGACACCGGAGCTCGC 1116  
Qy 977 CCGTCCCGGAGAGCAATATCTCGGCGCCATCCATGAGCAACGAGAAATAAGAGCGGCA 1036  
Db 1117 CCGGCTC---GACGACACCGTCCGCAACGTGACGCTTTCACCAACGAGAGGGCA 1172  
Qy 1037 GCCCGTTGGCGAGAGAGCGGCGCTGCGCTGCGCTGCGCTCCCGGACATGTTCCGGA 1096  
Db 1173 ACTGATCGGCGCGCGGAG--CGAGTCCCGCTTGGCGCTGTGGCGCTCCCGGACTTTC 1231  
Qy 1097 AGCATCTCGCGCGGCTGACGCGCTGTACGCGCAAGCCCATCTACATCAAGAGACGAT 1156  
Db 1232 CTGGTGTGATACAGCAAGGTAC---CTACCGCGCATCTACGTACAGGAACGGCA 1287  
Qy 1157 GCCCGTCCCTGAGAGAGGAACAATGACGTGCGAGAGGCGGTCAACGACCCCTTCCGA 1216  
Db 1288 CGAGCATCAAGGCGAGAGGAGCTTGGCCAAAGAGAAATTTCTGAAGATGACTTCAGGG 1347

Qy	1217	TCGCTACTTGACTCCGACTTGACTGATTTCCAGGCCATTACCAGACGGCGCTCG	1276
Db	1348	TCAGTACTATTAAGATATACATCCGTCATGTTATACCGCGTGGACTGACGGGGTCA	1407
Qy	1277	TCGTCAGGGGATCTTTGGCGGCTTGGCTCGATTAACCTTGAATGTCAGATGGCTACG	1336
Db	1408	ACGTCAAGGGGATCTTTGGCGGCTTGGCTCGATTAACCTTGAATGTCAGATGGCTACG	1467
Qy	1337	GACCGATTCGGCGTACGCTTCAAGACTA	1367
Db	1468	TGACGAGTTGGGATTCATGATGATTA	1498
RESULT 6			
AK119546			
LOCUS	AK119546	1828 bp	mRNA
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:001-207-A05, full insert sequence.	linear	PLN 29-OCT-2003
ACCESSION	AK119546		
VERSION	AK119546.1		
KEYWORDS	FLI CDNA; oligo capping.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE			
AUTHORS	1 The Rice Full-length CDNA Consortium, National Institute of Agricultural Sciences Rice Full-length CDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otsuno, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurotsaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, A., Kondo, S., Konno, H., Miyazaki, A., Otsu, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.		
TITLE	Collection, mapping, and annotation of over 28,000 CDNA clones from japonica rice		
JOURNAL	Science	301 (5631), 376-379	(2003)
MEDLINE	22752273		
PUBMED	12869764		
REFERENCE			
AUTHORS	2 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashidume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Horii, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, A., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishikura, J., Nomura, K., Numasaki, R., Ohneda, E., Ohta, Y., Ohtsuki, K., Oka, M., Ooka, H., Otsu, N., Ota, Y., Otsuno, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, K., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaki-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yanagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and Yoshimura, A.		
TITLE	Collection, mapping, and annotation of 28k full-length cDNA clones from japonica rice		

JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1828)
AUTHORS	Kikuchi, S.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT	This clone is one of the 32k full-length cDNA clones from japonica rice. URL: <a href="http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_CDNA_Project_Team_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yanagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.,Yamamoto,M.,andNakahama,Y.FAIS_Genome_Sequencing_&amp;_Analysis_Group:_Otsuno,Y.,Iida,Y.,Fujimura,T.,Ikedo,R.,Ishibiki,Y.,Kawamata,M.,Kobayashi,M.,Kodama,T.,Kurotsuki,T.,Kusumegi,T.,Lu,M.,Masuda,H.,Miura,J.,Mizuno,K.,Narikawa,R.,Nikura,J.,Oka,M.,Ryu,R.,Sugano,S.,Sugiyama,A.,Suzuki,Y.,Tsunoda,Y.,Ueda,M.,Xie,Q.,Yokomizo,S.,Yoshimura,A.,Matsubara,K.andMurakami,K.Genome_Exploration_Research_Group_in_Riken:_Adachi,J.,Aizawa,K.,Akimura,T.,Arakawa,T.,Carninci,P.,Fukuda,S.,Hanagaki,T.,Hara,A.,Hashizume,W.,Hayashida,K.,Hayatsu,N.,Hiramoto,K.,Hirooka,T.,Hori,F.,Iida,J.,Imamura,K.,Imotani,K.,Ishii,Y.,Itoh,M.,Kagawa,I.,Kanagawa,S.,Katoh,H.,Kawai,J.,Kishikawa-Hirozane,T.,Kojima,Y.,Kondo,S.,Konno,H.,Kouda,M.,Koya,S.,Kurihara,C.,Matsuyama,T.,Miyazaki,A.,Murata,M.,Nakamura,M.,Nishi,K.,Nomura,K.,Numasaki,R.,Ohno,M.,Osato,N.,Ota,Y.,Saitoh,H.,Sakai,C.,Sakai,K.,Sakazume,N.,Sano,H.,Sasaki,D.,Sato,K.,Shibata,K.,Shinagawa,A.,Shiraki,T.,Sogabe,Y.,Tagami,M.,Tagami-Takeda,Y.,Tagawa,A.,Takahashi,F.,Takaku-Akahira,S.,Tanaka,T.,Tomaru,A.,Toya,T.,Waki,K.,Yasunishi,A.andHayashizaki,Y.Location/Qualifiers">http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-length_CDNA_Project_Team_Kikuchi,S.,Satoh,K.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yanagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T.,Yamamoto,M.,andNakahama,Y.FAIS_Genome_Sequencing_&amp;_Analysis_Group:_Otsuno,Y.,Iida,Y.,Fujimura,T.,Ikedo,R.,Ishibiki,Y.,Kawamata,M.,Kobayashi,M.,Kodama,T.,Kurotsuki,T.,Kusumegi,T.,Lu,M.,Masuda,H.,Miura,J.,Mizuno,K.,Narikawa,R.,Nikura,J.,Oka,M.,Ryu,R.,Sugano,S.,Sugiyama,A.,Suzuki,Y.,Tsunoda,Y.,Ueda,M.,Xie,Q.,Yokomizo,S.,Yoshimura,A.,Matsubara,K.andMurakami,K.Genome_Exploration_Research_Group_in_Riken:_Adachi,J.,Aizawa,K.,Akimura,T.,Arakawa,T.,Carninci,P.,Fukuda,S.,Hanagaki,T.,Hara,A.,Hashizume,W.,Hayashida,K.,Hayatsu,N.,Hiramoto,K.,Hirooka,T.,Hori,F.,Iida,J.,Imamura,K.,Imotani,K.,Ishii,Y.,Itoh,M.,Kagawa,I.,Kanagawa,S.,Katoh,H.,Kawai,J.,Kishikawa-Hirozane,T.,Kojima,Y.,Kondo,S.,Konno,H.,Kouda,M.,Koya,S.,Kurihara,C.,Matsuyama,T.,Miyazaki,A.,Murata,M.,Nakamura,M.,Nishi,K.,Nomura,K.,Numasaki,R.,Ohno,M.,Osato,N.,Ota,Y.,Saitoh,H.,Sakai,C.,Sakai,K.,Sakazume,N.,Sano,H.,Sasaki,D.,Sato,K.,Shibata,K.,Shinagawa,A.,Shiraki,T.,Sogabe,Y.,Tagami,M.,Tagami-Takeda,Y.,Tagawa,A.,Takahashi,F.,Takaku-Akahira,S.,Tanaka,T.,Tomaru,A.,Toya,T.,Waki,K.,Yasunishi,A.andHayashizaki,Y.</a> Location/Qualifiers
FEATURES	1..1828
SOURCE	/organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /clone="001-207-A05"
ORIGIN	
Query Match	19.1%; Score 278.2; DB 8; Length 1828;
Best Local Similarity	54.7%; Pred. No. 4.1e-42;
Matches	785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;
Qy	20 TGCACAACGACTTGAATGGGCTTCCCAACGCGCGCTACAGATCGAAGCGCCGCTCA 79
Db	208 TCCCGAGGGGTTCGTTGGCGACCGCTCCCGGTACGATGAGAGGACTGTGA 267
Qy	80 AAGAAGTGGCCCGCGCCCTTCATCTGGGACAGTACGTCGACCTGGAGCCATCGCGCA 139
Db	268 AGAGGACGGGAGGCGACACCATCTGGGACAGCTTGGCGCACACTTTGGAAGATCA 327
Qy	140 CCAAGCGCCCAACGCGGATGTGGCTTGGCATCTACACCGCTAGATGAGGACTTTG 199
Db	328 CGGACTTACCAAGCTGATGTTGCAAGTATGATGATGATGATGATGATGATGATGAT 387
Qy	200 ATCTCTTACCAAGTACGCGGCAAGGCTTACCGCTTCTCTTGTGTGTCGGCATCA 259
Db	388 AGCTCATGGGACATGGGATGATGCGTATCGATTCGATGATGATGATGATGATGATGAT 447
Qy	260 TTCCTTCGCGCGGAGCTGATCCGTCACAGAGAGGAAATTGATTTTACGCAAC 319
Db	448 ACCCAATGCG-----TTGTCAAGTCAATCAAGCTGGATGACCACTTCAACAGC 501
Qy	320 TGAATGAGCCCTGTGAGGCGGGGATACAGCTTGGGTGACTTTGTACCACTGGGATC 379

Db 502 TGATGATGCACTTTAGCAAAAGAAATTCAGCCATATGTGACTTACCACTGAGACC 561  
 Oy 380 TGCCTCAGGCGCTTCACGATGCTATGAGAGCTGCTCAACGTGAAAGGTCACGCTGG 439  
 Db 552 TCCCCAGGCGCTTTAAGCAAGTACAAAGGCTGCTGACAGGCA---GATAGTGAACG 618  
 Oy 440 ACTTTGACCGGTATGCGAGGTGTGCTTTGAACTGTTGGGAGACGAGTCCAGAACTGGA 499  
 Db 619 ATTTGCGGCGTTACGCGGAGAGCTGCTTCAAGGAATTTGGGGACAGGGTGAAGCACTGGA 678  
 Oy 500 TCACCATACAGAACGCTTCGATTCAGGCTATCTATGATATGATCCACGCGGACCAACGCCC 559  
 Db 679 TCACGCTTACGAGCGGACACAGGTGGCTATCAGGGCTTACGACGCGAGGCTTCCAGGCGCC 738  
 Oy 560 CGGCGAGAGGAGC-----ATTAAACAAGCACTCCACGAGGGGCAACCTGCACTGAGC 613  
 Db 739 CCGGCGGCTGCTCCTGCTGCTCCACTCTATGTAAGGCGGCACTTCGCGACCGAGC 798  
 Oy 614 GGTGCTGCTGTAAGAAAGGCGGACATGATGAGCCATGCGCGCGGTGGCGCTTACAGCA 673  
 Db 799 CCTACGTGTCGCCACCATTTCACTTCGCGCCAGCGCGCGCGCGCATCTTACAGGA 858  
 Oy 674 GGGGCTTTGCGCCCTCGGAAAGGCGGACATCGGATCTGCTCAAGCGGCACTACTATG 733  
 Db 859 CAAAATACAGGCGACGAGAAAGCGGCGCTTGGATAGCGTTGACGTGATGTGTTGCG 918  
 Oy 734 AGCCCTGAGACAGCATGAGCTCGGACCAAGAGGCTGTGAGCGAGCGATGGAATTTTC 793  
 Db 919 AGCGGATGTCCAAACCAACGATC---GACATCGAGGCGCGCAAGAGCGAGAGTTTC 975  
 Oy 794 ACATTGGCTGTTTGGCCATCCATCTTCTTGAAGAAAGACTATCCAGAGAGCATGAGA 853  
 Db 976 AGCTAGAGATGTTGCTGATCCGTTCTTCTT---CGGCACTACCGCGCGACGATGAGGG 1032  
 Oy 854 AGCAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGGAATTGGCCATCTCTCAATCCG 913  
 Db 1033 CGAGGCTGGGAGAGGCTGCGGAGTTCAAGCGGATGAGGCGCGCGCTGCTCAAG---G 1089  
 Oy 914 GAGAGACGACTTCTTACGCGATGAATTAACAATCCAGTTGCGCGGCACTTACAGC 973  
 Db 1090 GGGCGCTGATTTGCTGGGCAATAACCACTACCACTACTACAGGAGCAACAAACA 1149  
 Oy 974 GTCCGCTCCCGA---CGGACTATCTGGGCGCATCTCATAGCAACGAGAAAT-- 1026  
 Db 1150 CCAACATCATCGGACATGTGCTCAACAAACCTTGGAGACACCGGACCGCTGACCTCC 1209  
 Oy 1027 ---AAGACGAGCGCGCGCTTGGCGAGAGAGCGGCTTCGCTGGTGGCTCCG 1081  
 Db 1210 CATTCAAGAAATGGAGAGCAATTTGAGATAGGCAAAATTCGATATGCTGTACATTGTGC 1269  
 Oy 1082 CGGACATGTTCCGGAAGCATCTCGCCGGGTGTAAGCGCTGT--ACGGCAAGCCCATCT 1138  
 Db 1270 CCCGAGGATGAGGAGCGCTGTAATATGTCAAGAAAGGTACACAGCCCACTGT 1229  
 Oy 1139 ACATCACGAGAAAGGATGCGCCGCTGCGCTGAGAGAGAAACATGACGTGAGAGAGCCG 1198  
 Db 1330 ACATCATCTGAAAGCGGATGAGATGACAGCAACACCGCTTCAATTCATCAAGAGACGCC 1389  
 Oy 1199 TCAACGACCCCTTCCGATCGSTACTTGTACTGCACTTGCATTCGATTTCCAAAGSCCA 1258  
 Db 1390 TCAAGGACAGAGAGATCAAAATACCAATGATCTACCAACATCTGGCTGCTTCCA 1449  
 Oy 1259 TTAACCGAGAGCGGCTGTGTCAAGGGGTACTTTGCTGGGCGGTGGCTGATTAATTGG 1318  
 Db 1450 TCAAGGAGAGCGGCTGTGAGCTAGCTGTGATCTTGGGTCTGTGAGCAACTGGG 1509  
 Oy 1319 AATGCTCAGATGGCTACGAGCCAGATTCGCGCGTCACTTCAAGACTA---CAACC 1375  
 Db 1510 AGTGGGCGCGGATCTCTCGAGATTCGCGCTTCACTTGTGAGATTAAGAGATTAAC 1569  
 Oy 1376 TCAAGCGACGCGCAAGAGTCTGCTGCTCAAGAGCATGTTGCGGCGG 1430  
 Db 1570 TCAAGAGATACCCCAAGAACTCGGTGAGTGTCAAGGCTCTCTGAACACTG 1624

RESULT 7  
 AY129294  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzoae; Oryza.  
 1 (bases 1 to 2032)  
 Akiyama, T., Opassiri, R., Ketudac-Cairns, J. R., Svast, J., and Bsen, A.  
 Predominant expression of beta-glucosidase in germinating rice  
 seeds  
 JOURNAL  
 TITLE  
 AUTHORS  
 DIRECT SUBMISSION  
 JOURNAL  
 Submitted (05-JUL-2002) Low Temperature Science, National Hokkaido  
 Agricultural Research Center, 1 Hitsujigoka, Toyohira-ku, Sapporo,  
 Hokkaido 062-8555, Japan  
 Location/Qualifiers  
 1..2032  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Yukinikari"  
 /db\_xref="taxon:39947"  
 142..1707  
 /codon\_start=1  
 /product="beta-glucosidase"  
 /protein\_id="AA01354.1"  
 /db\_xref="GI:22658440"  
 /translation="MGRISSSGSGSTARLEAVAVLVVGVVSSLRGCIAGQSGG  
 LTRGSPFEGVGTSAAYQYGAVEDRGGTITMDTPHTEGKLTDFENAVADQY  
 HRPEDIOIMADMGMQDAVRFSTAMSRIVNGYQUNQAGITDYNKLTIDLLAKGTQY  
 VTLIYHNDLQALBEDRYKGLDROIYDDPAVAIETCFREGRDVKHMTILNEHTVAIQ  
 GHDAGIQAFGRCSVLHLIYCKAGNSGTEPYVAHHFLTLAAHAASIRYKRYATQNG  
 LGIAFDVVMKGLDFVGINHTYTYRHNNTNIGTLNNTLADGTGVSLEPFGKQDYG  
 TADEAVVNGALDFVGINHTYTYRHNNTNIGTLNNTLADGTGVSLEPFGKQDYG  
 DRANSIWIYIVRGRMSLMNYKERNISPVYITENGMDSNPFISIDALKDSKRI  
 KYHNDYLTNLAAISIKEDGCDVAGRFAMSLDLMWMAAGVSSRFLYFVDYKDKLRYP  
 KNSVQMFRLTKT"  
 ORIGIN  
 Query Match 19.1%; Score 278.2; DB 8; Length 2032;  
 Best Local Similarity 54.7%; Pred. No. 4e-42;  
 Matches 785; Conservative 0; Mismatches 608; Indels 42; Gaps 10;  
 Oy 20 TGGCCACGACCTTGAATGGGCTTGGCAAGCGCGGCTTACCAAGATCGAAGCGCGCTGA 79  
 Db 290 TCCCGAGGGGTTGCTTCGACACCGCTTCGCGCGGCTTACGAGGAGGCTGTGA 349  
 Oy 80 AAGAAAGTGGCGCGGCGCGCTTCATCTGGGACAGCACTGCGACCTCGAGCGCATGGCGCA 139  
 Db 350 AGGAGAGCGGAGAGGCGGACCACTTGGGACAGCTTGGCGACACCTTTGGAAGAGATCA 409  
 Oy 140 CCAACGCGCCAAAGCGGATGTGGCTTGGATCACTACCAACCGCTACAGATGAGAGACTTTG 199  
 Db 410 CGGACTTCAGCAATGTGATGTTGACAGTTGATCACTACACCGTTTTCGAGGAGGATATAC 469  
 Oy 200 ATCTCTTGAACAAAGTACGCGCAAGGCGCTTACCGCTTCTCTTGTGCTGTGGCGATCA 259  
 Db 470 ACTCATGCGACAGATGGGAGTGTATCGGTTCGATACCAATGTGTCAAAATCT 529  
 Oy 260 TTCCCTCGGCGGCGAGGCTGATCCGTCACAGAGAGGAGATGAGTTTACAGCAAC 319  
 Db 530 ACCCAATGTG---TTGCTCAAGTCAATCAAGCTGTGTGACCACTACCAACAGC 583

QY 320 TGATTTGACGCGCTTTGAGCGGGGTATTCAGCCTTGGTGATCTTGTATACCATGGGATC 379  
DB 584 TGATCGATGCACTTCTAGCAAAAGGAATTCAGCAATATGTGACACTTACACATGGGAGC 643  
QY 380 TGGCTCAGGCGCTTCAGCATCGCTATGAGGCTGGCTCAACGTGGAAAGAGTCCAGCTGG 439  
DB 644 TCCCCAGGCGCTTCGAAAGACAAAGGAGTGGCTTGAACAGCA--GATATGTGACG 700  
QY 440 ACTTTGAGCGGTATGCGAGGTGTGCTTGAACGTTTGGGGACCGAGTCCAGAACTTGA 499  
DB 701 ATTTCGGGGGTAGCGGAGACGCTTCAAGGAGTTCGAGGAGGAGGATGAAGCACTTGA 760  
QY 500 TCACCATCAACGAWCCCTGATTCAGGCACTTATGATATGACCCGCGAGCAACGCC 559  
DB 761 TCAGGCTCAACGAGCGGACACGCTGGCCATTCAGGGCTACGAGCGAGGCTCCAGGCC 820  
QY 560 CGGGCAGGAGCAGC-----ATTAAACAAGCACTCCACCGAGGCAACATGCGCATGAGC 613  
DB 821 CCGGCGCGCTCTCGTCTGCTGCTCAGCTTACTCAAGGCGGCACTCCGCGCACCGAGC 880  
QY 614 CGTGGCTGCTGGAAGGCCAGATTCATGAGCCATGCGCGCGCGCTGAGCGCTTACAGCA 673  
DB 881 CTTACGCTGTCGCCACACTTCTCTCGCCACGCGCGCGCGCGAGCATTCACAGGA 940  
QY 674 GGGACTTTCGCGCTCGCAAAAGGCGCAGATCGCATCTGCTCAACGCGCATCTATG 733  
DB 941 CAAATTCAGAGGCGACCCAGACCGGAGCTTGGGATAGCTTGAAGTATGTGCTTCG 1000  
QY 734 AGCCCTTGGGACAGCAATGAGCTTCGGGACAGAGGCTGTCAGCAAGATGGAATTC 793  
DB 1001 AGCCGATGTCCACACACACGATC---GACATCGAGGCGGCGCAAGAGCGAGGATTC 1057  
QY 794 ACATGGCTGGTTTGGCAATCCCATCTTCTTGAAGAAAGATATCCAGAGAGCATGAAGA 853  
DB 1058 AGCTAGGATGGTTTGTGATCGCTTCTTCTT---CGCGATACCTCGCGCAGATGAGGG 1114  
QY 854 AGCAGCTGGGCGAGAGGCTTCAGACCCCTCACTCCGCGGCACTTGGCATCTCTAATGCG 913  
DB 1115 CGAGGCTGGGAGAGGCTTCGCGAGTTTACAGCGGCGATGAGAGCGCGCTCTCAAG---G 1171  
QY 914 GAGAGACGCACTTCTACGGCATGAATTAACATCCAGATTGCGCGGCCACTTAAGC 973  
DB 1172 GGGGCTGTGATTTGCTGGGCAATAACACTACACACTACTACAGAGGCAACAACA 1231  
QY 974 GTCCCGTCCCGAGA-----CGGACTATCTGCGGCGCATCATAGCAACGAGGAT-- 1026  
DB 1232 CCAACATCATCTGGGACATGTGCTCAACAACACTTGGGACAGACCGGCACTGCTGCC 1291  
QY 1027 -----AAGACGCGGAGCGCGCTTGGGAGAGAGCGGCGCTGCGCTGGCTGCGCTGCC 1081  
DB 1292 CATTCAAGAAATGGGAAGCCAAATTTGGAATAGGCGAAATTCGATATGGCTGTACATTGTC 1351  
QY 1082 CGGACATGTTCCCGAGACATCTCGCCCGGGGTGAACGCGCTGT---ACGGCAAGCCCATCT 1138  
DB 1352 CCGGAGGATGAGAGGCTGATGAATATGCAAGAAAGGTACAAACAGCCCAAGTGT 1411  
QY 1139 ACATCAGCGAAGAGGATGCGCGCGCTGCGAGAGGAAACATACGTCGAGAGAGCGCG 1198  
DB 1412 ACATCATGTGAAGAGGAGTGAAGACAGCAACCCGTTCTATTTCATCAAGAGACGCC 1471  
QY 1199 TCACGAGACCCCTTCCGATCGGTAATCTTGAATGCACTGCACTGCAATTCGAAGGCCA 1258  
DB 1472 TCAGAGGACAGCAAGAGATCAAAATACCAATATGACTACCAATCTGGCTGCTTCCA 1511  
QY 1259 TTACCCAGAGCGGCTGTGCTCAAGGGGTACTTTCGCTGGGCGCTTCTCGATTAATTCG 1318  
DB 1532 TCAGAGGAGAGCGGCTGCGAGCTACGTGAGTACTTTCGCTGGTCTGCTGCAACTGGG 1591  
QY 1319 AATGCTCAGATGCTTACGAGACCCAGATTCGCGCTCAGCTTACAGACTACG---CAACC 1375  
DB 1592 AGTGGGCGGCGGATCTCTCGAGATTCGGGCTCTACTTCTGTGACTACAGAGATTAAC 1651  
QY 1376 TCAAGCGACGCGCAAGAGTGTGCTGCTCAAGAGCATGTTTGGCGGCCG 1430

DB 1652 TCAGAGATACCCCAAGAACTCGTGCAGTGTTCAGAGCGCTCTGAAAGCTCG 1706  
RESULT 8  
AF268911  
LOCUS  
DEFINITION  
Aspergillus niger beta-glucosidase precursor (bg11) gene, complete cds.  
ACCESSION  
AF268911  
VERSION  
AF268911.2  
KEYWORDS  
GI:8572535  
ORGANISM  
Aspergillus niger  
Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE  
1 (bases 1 to 1988)  
Al,Y.-C. and Meng,F.-M.  
Molecular cloning and sequencing of a beta-glucosidase gene from  
Aspergillus niger AMS11  
Unpublished  
JOURNAL  
2 (bases 1 to 1988)  
Al,Y.-C. and Meng,F.-M.  
Direct Submision  
TITLE  
Submitted (17-MAY-2000) Department of Biochemistry, Zhongshan  
University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.  
China  
REFERENCE  
3 (bases 1 to 1988)  
Al,Y.-C. and Meng,F.-M.  
Direct Submision  
TITLE  
Submitted (19-JUN-2000) Department of Biochemistry, Zhongshan  
University, 135 Xingang Road, Guangzhou, Guangdong 510275, P. R.  
China  
REMARK  
Sequence update by submitter  
On Jun 19, 2000 this sequence version replaced gi:8248474.  
COMMENT  
Location/Qualifiers  
FEATURES  
1. 1988  
/organism="Aspergillus niger"  
/mol\_type="genomic DNA"  
/strain="AMS11"  
/db\_xref="taxon:5061"  
<496..>1858  
/gene="bg11"  
join(<496..536,595..>1858)  
/gene="bg11"  
/product="beta-glucosidase precursor"  
join(496..536,595..1858)  
/gene="bg11"  
/codon\_start=1  
/product="beta-glucosidase precursor"  
/protein\_id="AA074209.2"  
/db\_xref="GI:8572535"  
/translation="MLPQLQWGFAPAAVQIEGAVDQDGRPSIMDTFCAOPGKIADG  
SSGVTAQDSYNTAEADIALKSLGAKSYRPSISRIPEGGRDAVAVQAGIDHYVKEVD  
DLIDAGITPTITLPHMDLHORYGALNTEPLDENYARVAFRALPKVRNNEPLC  
SAIRYGGSGFAPRGROSTSEPMVGHNIIVAHGRAYKAVYDDPKPSAGDQIGLVNG  
DPTFPMADADPAKRELAEPTAPFADPIYGVDPAYAMRQYLDGRLPFTPEEPAIYAG  
SNDPFGNHTSYIIRHRSPPASADTGVAVDVLFTNKGQNCIGPETSQWLPKCAAG  
FRDPLVTSKRYGSPPIYVTEGTSIKGSDDLNEKILIEDDFVKKYNEYIRAMVAV  
ELDGVAVRFGVTVVDYENGQKRPKKSASLKLPLBELIAAA"  
sig\_peptide  
mat\_peptide  
ORIGIN  
Query Match 17.4%; Score 253.2; DB 8; Length 1988;  
Best Local Similarity 56.0%; Pred. No. 1.9e-37;  
Matches 708; Conservative 1; Mismatches 474; Indels 81; Gaps 9;  
QY 45 CGCAACGCGCGCTACCATCGAAGCGCGCTCAAGAAAGATGCGCGCGCGCTTCAT 104

Db 583 CCCGTCGACAGCTACCAAGATGAGGGCGCGTGCAGCAAGACGGCGCGGCCAGCAT 642  
 Qy 105 CTGGAGACAGTACTGCGACCTGGAGCCATTCGGCGACCAACGGCGCCAAAGCGATGTGAC 164  
 Db 643 CTGGAGACAGTACTGCGCGAGCCCGGGAAGATTCGCCACCGCTGTGGCGCGTGAACGGC 702  
 Qy 165 TTGGCATCTACTACACCGCTACGATGAGAGCTTTGATCTTTGACCAAGTACGGCGCAA 224  
 Db 703 GTGCACCTGTAAACACCGCACGGCGAGAGACATTCGCTCAAGTGCCTCGGGGGCCAA 762  
 Qy 225 GACCTACCGCTTCTCTGTGTGGTGGGGATCATTCCTCCGCGGGGAGCGTGGATCC 284  
 Db 763 GAGCTACCGCTTCTCTCATCTGT-----CGCGATCCCGAGGGCGCGCGCGATGC 816  
 Qy 285 CGTCAACGAGGAGGGAATTGATTTTACAGCAAACTGATGACCGCTTTGAGCGGGG 344  
 Db 817 CGTCAACGAGGCGGCGATCGACCACTACAGTTGCTCGACGACCTGCTCGACCGCG 876  
 Qy 345 TATCAGCGCTTGGGATCTTTGATCTGGGATTCGCTCAGCGGCTTCAAGATCGCTA 404  
 Db 877 CATCAGCGCTTCACTACCGCTTTCACCTGGGACCT-----GCTGCACTCAGCGGTA 927  
 Qy 405 TGAAGCTGGCTCAACGTGAAAGAGGTCCAGCTGGACTTTGAGCGGTATGGAGGTTGTG 464  
 Db 928 CGGGGGGCTGCTGAACCGCACCGAGTTCCGCTGACTTTGAAAATAAGCCCGCTCAT 987  
 Qy 465 CTTTGAACGTTTGGGAGCAGATTCAGAACTGATCAACATCAACGAMCCCTGATTTCA 524  
 Db 988 GTT-----CAGGGCGGTGCCAAGGTGGCGAACTGGAACGAGCGCGTGTCTC 1035  
 Qy 525 GGGCATCTATGATATGCGACCGCGAGCAAGCCCGGGCGAGAGCAGCATTAACAGCA 584  
 Db 1036 GGCATTCGGGAGCTACGGCTCCGGAGTGTGCCCCCGCGCGAGACCA----- 1085  
 Qy 585 CTCACCGAGGGCAACACTGCACTGAGCCGTGCTGCTGAAAGGCCCAAGATCATGAG 644  
 Db 1086 -----CCTCGAGCCGTGAGACGCTCGGCCAACAATCTCTGTGC 1125  
 Qy 645 CCATGCCCGCGCGCTGCGCTTCAAGAGGAGACT---TGCCTTCGCAAAAGGGCCA 701  
 Db 1126 CCAGCGCGCGCGCTCAAGGCGTACCGCGAGCACTTCAAGCCCGCAGCGCGAGCGCA 1185  
 Qy 702 GATGGGATCTTCGCTCAACCGCGACTACTATGAGCGCTTGGAGCAAGCATGAGCTCGGGA 761  
 Db 1186 GATGGGATCTTCGCTCAACCGCGACTACTATGAGCGCTTGGAGCAAGCATGAGCTCGGGA 1245  
 Qy 762 CAAGAGGCTGCTGAGCAGCGATGAGATTTCACTTGGCTGTGTTTGCATTCGCATCTT 821  
 Db 1246 CAAGAG-----CGCTCGAGTTCTTCAAGCGCTGTGTTTGCAGATTCCTCATTTA 1293  
 Qy 822 CTGGAAGAGGACTATTCAGAGAGAGATGAGAAGAGCTGGGCGAGAGGCTTCCAGCCCT 881  
 Db 1294 CTTGGGC---GACTTACCGCGCGTGCATCGCAAGAGCTGGGCGAGCGGCTGCCGACTT 1350  
 Qy 882 CACTCCCGCGGACTTTGCGATCTTCAATGCCGAGAGCCGACTTTTACCGCATGAATTA 941  
 Db 1351 TACGCCGAGGAGCGCGCTCTGTCA---CGGCTCCAAACGACTTTTACCGCATGAACCA 1407  
 Qy 942 CTACACATCCAGTTTCGCGCGCCACTGACGATTCGCCGCCGAGAGCGACTTATCTCGG 1001  
 Db 1408 CTACACATCCAGTTTCGCGCGCCACTGACGATTCGCCGCCGAGAGCGACTTATCTCGG 1467  
 Qy 1002 CGGCATTCATGAGCAACGAGGAATGAGAGCGGAGCGCGCTGGGCGAGAGCGCGCT 1061  
 Db 1468 CAACGTGACGCTGCTTCAACCAACAGGAGGCAATGATCGGCCCGCGAGAGCGCATGC 1527  
 Qy 1062 CGCGTGGCTGCGCTTCGCGCGGAGCATTTCCGGAAGATTCGCCCGGGGTATACGCGCT 1121  
 Db 1528 CCGCTGGCTGGCGCTCTGTGCGCGGCTTCCGCACTTCTGTGTGTGAGCAACAGAG 1587  
 Qy 1122 GTACG---CAAGCCATCTATCATCAACGAGAGCGATGCGCGCTTCGAGAGGAGAA 1178  
 Db 1588 GTACGCTCCCGCCCATCTATGATGAGAGAGACCGCACGAGCATCAAGGGCGAGAGCGA 1647

Qy 1179 CATGACGTGAGAGGCGCGTCAACGACCCCTTCCGATCCGCTTACTTGTGACTCGACTT 1238  
 Db 1648 CTTGCCAAGAGAAAGTTCTCGAAGATGACTTCAAGGCTCAAGTACTATTAACGATCAT 1707  
 Qy 1239 GGACTGATTTCCAGGCGCTTACCCAGAGACGCGCTGCTGTCAAGGGATCTTGGCTG 1298  
 Db 1708 CCGTGCATGTTTACCCCGTGGAGCTGAGACGGGGTCAACGTGAGTTTGGGGTTACGTA 1767  
 Qy 1299 GCGG 1302  
 Db 1768 TGTG 1771  
 RESULT 9  
 AK067001  
 LOCUS  
 DEFINITION  
 Oryza sativa (japonica cultivar-group) cDNA clone:J013092F09, full insert sequence.  
 ACCESSION  
 AK067001  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE  
 AUTHORS  
 The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ohtomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kuroseki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusneger, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN; Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishi, Y., Itoh, M., Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Otsu, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.  
 Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
 Science 301 (5631), 376-379 (2003)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 2045)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishi, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Kuroseki, T., Kusneger, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Otsu, N., Ota, Y., Ohtomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takehashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Taya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.  
 Direct Submission

TITLE

## JOURNAL

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kennondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: shikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL: <http://cdna01.dna.affrc.go.jp/cDNA/>

NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Negata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yarakhi, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otsuno, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurobaki, T., Kusumegi, T., Lu, M., Masuda, H., Mura, T., Mizuno, K., Naitkawa, R., Naitkawa, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken: Adachi, J., Aizawa, K., and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Haseizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Otsu, N., Oda, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takata-Akashira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

## FEATURES

## Source

1. 2045  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J013092F09"

## ORIGIN

Query Match 17.1%; Score 248.6; DB 8; Length 2045;  
Best Local Similarity 52.3%; Pred. No. 1.4e-36;  
Matches 730; Conservative 1; Mismatches 632; Indels 33; Gaps 7;

20 TGGCCACGACTTGAATGGGGCTTGCACAGCGCCGCTTACCAAGTGAAGCGCCGTCA 79  
Db TCCCGAGGAGCTTCTCTCGGCAACCGGCTCCGGGCTTATCAGTACGAGGCGCTGTGA 315  
QY AAAAAAGTGGCGCGCGCCCTCATCTGGGACAGTCTGCGACCTGAGACCATCGGCA 139  
Db AAAAAAGTGGCGCGCGCCCTCATCTGGGACAGTCTGCGACCATCTGAGGCGCA 375  
QY 140 CCAAGCGCGCAAGCGGATGTGGCTTGCATCTCAACCGCTACAGTGAAGGACTTTG 199  
Db 376 AAGATGTAGCAACGGGAGATGTAGCACTTTTACATCGCTACAGAGAGATTTGA 435  
QY 200 ATCTCTTGACCAATAGCGGCAAGGCGCTTCTCTTGTCTGTGTGCGGATCA 259  
Db 436 ACTTCTGACCGCATGAATGAGACGCTTCCGTTTCTCATTTGAGAGCAGATTC 495  
QY 260 TTCCCTCGGCGGAGGCTGATCCGCTCAACGAGGAGGATTTGAATTTCAGCAAC 319  
Db 496 TGGCAATGAAACCATCATGAGGAGATCAACAAAGGGGATTTGCTTTTCAACAGCC 555  
QY 320 TGAATGACGCTTGTGAGCGGCGGTATCAAGCTTGGTGAATTTTGTACCATGAGATC 379  
Db 556 TGAATTAACGAGTCAATCTAGAGGTTGAAGCATTTGTTCACATCTTTTCATTTGACA 615  
QY 380 TGGCTCAGGCGCTTACGATCGCTATGAGGCTGGCTCAAGTGAAGAGGCTCCAGCTGG 439  
Db 616 CCCACAGGCTTGAAGGACAAATATACCGGAGCTTCTCA---GTGAATAATTTGTGAAG 672

QY 440 ACTTTAGCGCGTATGCGAGGTTGTCTTGAACGTTTGGGAGCCAGTCCAGAACTGGA 499  
Db 673 ATTTCGTGACATATGGGACAGTGTGCTTCCGTAGTTCGGCGACCGGTTAACTCATGA 732  
QY 500 TCACCATCAACGAWCCCTGATTCAGGCATCTATGATATGTCACCGGACGAAAGCCC 559  
Db 733 ACAGTTCACAGAGCGGATGATTTCTTGGCGCGGCGGCTTACGCGCTCGGACCAAGCCC 792  
QY 560 CCGGACGAGGAG-----AGATTAAAGACATCCACGAGGCGAACATGCCACTGAGC 613  
Db 793 CCGGCGCTGTGCGCGTGTGCTTCAAGAAATGCGCCCGGCGAGCTCGGGAACGAGC 852  
QY 614 CTTGGCTGTGTAAGAGCCAGATCATGAGCCATGCGCGCGGCTGTGAGCTTACAGCA 673  
Db 853 CTTAGCTGCGCGGCGCAACACTGTCTGTGCGCCACCGGAGGCGCTTCTTACCGCC 912  
QY 674 GGAAGCTTTCGCGCCCTGCAAAAGAGGCGAGTGGGATCTTGTCTCAACGCGGACTTATG 733  
Db 913 AAAAGTACACGAGCGAGAGAGAGAGAGGAGATCGGCATCAAGGATGTGCACTGTTG 972  
QY 734 AGCCCTGGGACAGCATATGAGCTTCGGGACAAAGAGGCTGTGAGCGAGATGGAATTC 793  
Db 973 TGGCTT---ACAGCGACGCGCGCGCGGACAGAGAGCGCGGCGAGCTTGACTTCA 1029  
QY 794 ACATTTGCTGTGTTGGCAATCCCATCTTCTTGAAGAGAGCTATCCAGAGAGATGAGA 853  
Db 1030 TGTACGATGTGTTATGAGACCGATGTGT---CGGCGACTACCGGCGGACATAGAGA 1086  
QY 854 AGCAGCTGGGCGAGAGGCTTTCAGCCCTCACTCCGCGGACTTGTGCATCTCAATGCCG 913  
Db 1087 AGTTGGTCGGTGAACGGCTGCGCCCAATATTCACAGCGGAGAGTGGAGTTGTGA---GG 1143  
QY 914 GAGAGACCACTTCTACGGGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 961  
Db 1144 GATCTTACGACTTCTTACGCTGCTCACTCACTCACTCACTCACTCACTCACTCACTCA 1203  
QY 962 GCCACTTACGCGGCTGCGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021  
Db 1204 GAGCGCGGAG 1263  
QY 1022 AGAATTAAGACGAG 1081  
Db 1264 AGCGAAG 1323  
QY 1082 CGGACATGTTCCGGAAGCATCTGCGCGGAGTGAAGGCTGTACGCGGAGAGAGAGAG 1141  
Db 1324 CGGCTCTGCGTGAAGCTGCTTCTGTATACCAAGAGAGAGAGAGAGAGAGAGAGAG 1383  
QY 1142 TCACGAG 1201  
Db 1384 TTGACAG 1443  
QY 1202 ACAGACCTTTCGCGATCCGCTTCTTGAATCTGATCTGATCTGATCTGATCTGATCT 1261  
Db 1444 AGGATGACAAACGAAATATCTTCTCACTACAGAGATCTCGCT---TCACTCACTGCAA 1500  
QY 1262 CCAAG 1321  
Db 1501 TCAAG 1560  
QY 1322 GGTCAAGTGTGCGAG 1381  
Db 1561 GGGGCGAGCGCTACACCGGCGCTTGGGCTTCTTACAGTGTGAGCGGAACTCTCAAG 1620  
QY 1382 GCAAGCGCAAGAGATC 1397  
Db 1621 GTATGCAAGAGATC 1636

RESULT 10  
AF439322  
LOCUS AF439322 2068 bp DNA linear PLN 20-JUN-2002



	DEFINITION	Talaromyces emersonii beta-glucosidase 1 (bg1) gene, complete cds.
	ACCESSION	AJ439322
	VERSION	AJ439322.3
	KEYWORDS	GI:21493049
	SOURCE ORGANISM	. Talaromyces emersonii Talaromyces emersonii Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; Talaromyces.  
	REFERENCE AUTHORS TITLE	1 (bases 1 to 2068) Murray,P.G., Collins,C.M. and Tuohy,M.G. Molecular cloning and expression analysis of beta glucosidase from Talaromyces emersonii Unpublished
JOURNAL REFERENCE AUTHORS TITLE JOURNAL	2 (bases 1 to 2068) Murray,P.G., Collins,C.M. and Tuohy,M.G. Direct Submission Submitted (23-OCT-2001) Biochemistry, National University of Ireland, Galway, Ireland 3 (bases 1 to 2068) Murray,P.G., Collins,C.M. and Tuohy,M.G. Direct Submission Submitted (31-MAY-2002) Biochemistry, National University of Ireland, Galway, Ireland Sequence update by submitter 4 (bases 1 to 2068) Murray,P.G., Collins,C.M. and Tuohy,M.G. Direct Submission Submitted (20-JUN-2002) Biochemistry, National University of Ireland, Galway, Ireland Sequence update by submitter On Jun 20, 2002 this sequence version replaced gi:21284362. Location/Qualifiers 1..2068 /organism="Talaromyces emersonii" /mol_type="genomic DNA" /db_xref="taxon:68825" <304..>2068 /gene="bg1" join(<304..383,472..663,732..1204,1258..1843,1930..>2068) /gene="bg1" /product="beta-galactosidase 1" join(304..383,472..663,732..1204,1258..1843,1930..2068) /gene="bg1" /_EC_number="3.2.1.21" /note="glycosyl hydrolase" /codon_start=1 /product="beta-galactosidase 1" /protein_id="AAU34084_2"	
CDS	mRNA	/db_xref="gi:21284363" /translation="MTTEETPTTTTSITLPPPLMGFATASVIEGAFDEBGLPSIV DTFGPKVEDGTNGVDACSRYRTEDELEIKKGAKYRSLSMPRIIPLGAND PINEGLQFYKFLFDLLAAGISEPVTLPFMDLPDELIMKRYSGLMEKVADAVANYA RVNNAIGSKVKHWITNEPNKCISVLGHNTGKAAPRTSDRTSPEDGGREDAWIGH NLVAHGVNDTVIRREFRKOGEGITLTGNDAEPDNPENPADVCARCKIFBAISW PDPIHYHKIVDSNVKKQLGBRLPEFTFEELPAFHSGSDFKLMLADRNNNKLYVTENG PEDLAGINDILMEDKNAPIGETQGWELEFPUGEFKLFKWLDRNNNKLYVTENG TSVGESDKPLEEVLIINDEFROYQRDYTGAMVADVADDGVNAVMAWSLDNFEMSE GYSRFGVTYDVYKNGQRIPKSAIVIGELFNKYIKE"
ORIGIN	Query Match	17.0%; Score 247.8; DB 8; Length 2068;
	Best Local Similarity	53.2%; Pred. No. 2e-36;
Matches	Conservative	1; Mismatches 518; Indels 130; Gaps 5;
Df	472 CTACCAGATCGAAGGGGCCCTTCGACGAAGAATGCCTCCTTCGATTATGGACA	CTT
Ov	57 CTACCAACTCGAAGCCATCGCACCAAGAGGCGCCGCCCATCTGGGACAGTA	116
Df	117 CTGCAACTCGAAGCCATCGCACCAAGAGGCGCCGCCCATCTGGGACAGTA	176
Ov	532 CAGCAAAGCGCCCAGAACGTCGAGACGGAGCCAATGGCGATGTCGCTGCACTGTA	591
Df	177 CCACCGCTACGATGGAAGCTTTCCTTAGTCCATAAGTACGCGCAAGGCCTTAC	GCTT
Ov		236

Db	592	CCACCCTGACCGGTGAGACATTGAGATATTGAAGAAATACGGTGCACAGATCTATGCTT	651
Oy	237	CTCTCTGTCGTGCTGCG-----	252
Db	652	CTCGTTGTCTGTGTGTCTGTGTATGTCTGCTCTTCTTATTTTGTCTGATCTGTTT	711
Oy	253	-----CGGATCATTCCTCCCTGCGCGGACAGCTGATCCCGTC	288
Db	712	TGTCTGACAGAACAACTCCAGGCCACGATCATCTCTCTCGCGCGGCGCAAAAGCACCCATC	771
Oy	289	AACGAGAGAGGAATATGAGTTTATACAGAAATCTGATTAACCCCTCTGTGAGCGGGGATGC	348
Db	772	AATAAAAAGGGCCCGCAGTTCTACTCCAAAGTTCTTGATATGACTTACACGCCCGGGGATC	831
Oy	349	ACGCTTGGGATGACTTTGTACCACTGSGGATCTGCTCAGCGCTTCAACGATTCGATATGA	408
Db	832	GAGCCCTTCTGTCAAGCTCTTTCACCTGGGATCTGCGCGACGAGCTGATGAAAGCGGTACGG	891
Oy	409	GGCTGAGCTCAACGTGAAGAGGTCAAGCTGAGCTTTGAGCGGATATGCGAGTTGTCTTT	468
Db	892	GGCATGTCTCAACAGGAGGAGATTGTGGCGACTACGCCAACTATGCCCCGTGTCTTC	951
Oy	469	GAAGCTTTGGGGACCGAGTCCAGAACTGGATCCATCAACGATCCCTGGATTCAGGCC	528
Db	952	AACCGCTCTCGCTCGAAGGTCAACACTGTGATCTTTCAACGAGCCCTGTGCAATTCC	1011
Oy	529	ATCTATGGATATGCGACCGGCGAGCAAGCCCGGGGACGAGCAGACTTAAACAGCACTCC	588
Db	1012	GTGCTCGGCCCAACACCGGCGACAGCGCCCGGGGGGAGCAGCGATCCGACGAAAGT	1071
Oy	589	ACCGAGGGCAACACTGCGCACTGAGCCGTGTGCTGTGAAAAGGCCAGATCTAGAGCAT	648
Db	1072	CCGAGGGGAGACGGCACCCTCGGAACTTTGATCTGTGGGTCAACAACCTTCTGTGGCTCAT	1131
Oy	649	GCCCCGCGGTGCGCGCTTACAGCAGGACCTTTGCCCCCTGCAAAAGGCGCAATCGGC	708
Db	1132	GGTACCGTTGTGGACATCTATCGGAGAGAGTTCAAGAGAAACAAAGCGGGAAATCGGA	1191
Oy	709	ATCTGCGCTCAAC-----	720
Db	1192	ATACCCCTTCAACGGTAGGCTCGCGGCTCATCTCTGGAATTTGACTTATGTCTAACCGCATCT	1251
Oy	721	-----GGCGACTATATGAGCCCTGTGGACAGCAATGAGCCTCGGACACAGAGGCTGCTG	775
Db	1252	ATCTTAGGGGACTTGGGGCGGAACCATGTGGATCCGAAAAACCGACAGACGTCGAAACCTGCG	1311
Oy	776	AGCGACGATGGAATTTTCACTATGAGCTGTGTGTCGAATCCATCTTCTTGAAGAAAGACT	835
Db	1312	ACCGAAAGATGGAATTCGCACTCTGTGTGTGTGCGGAGCCCATCTAC---CACGGGAAT	1368
Oy	836	ATCCAGANAGATGAAGAAGCAGCGGGCGAGAGGCTTCCAGCCCTCACTCCCGGACT	895
Db	1369	ACCGGAGAGATGGTGAACAGTTTGGCGAGCGCTGTCCGCAAGTTCAACGCGGAAGAA	1428
Oy	896	TTGCATCTCTCAATGCGCGAGAGACCGCACTTCAACGCGATGAATTAATTACATCCAGT	955
Db	1429	TGCGCTTGTGTGACGGGACG---AAGCAATTTTAAGGATGAAACCACTACTGCGAAAACT	1485
Oy	956	TGCGCGCGCACCTTAGACGCTCCGCTCCCGAGACGGAATATCTGGCGCCATCCATGAGC	1015
Db	1486	ACATCCGCAACCGGACCGGCGAACCGAGACCGCGAGGACATTGTCTGGCAACTGGACATGC	1545
Oy	1016	ACGAGGAAATTAAGAGCGGACGCCCGTTGGCGAGAGAGGCGGCTCGCTGGCTGCGCT	1075
Db	1546	TGATGAGAGAACAGAACCGGCAACCCCATCGGCCCCGAGAACCCATGCAATGTGCTCGCG	1605
Oy	1076	CTGTGCGCGGACATGTTCCGGAAGACATCTGCGCGGGGTGTACGAGCTGTACGGCAAGCCC	1134
Db	1606	CGTTCCCGCTGGGGTTACGAAAGCTGCTGAAGTGGCTTGAGAACCGGATCAACAAACCCCA	1665
Oy	1135	--ATCTACATCAACGAGAACGATGCCGTGCTCTGAGAGAGAACATGACGTGCGAGG	1192



Db 1666 AGATCTACGTGACCGAGAACGGGACGCTCAAGGGCGAGAGCCGCTCGAGG 1725  
Qy 1193 AGCGCGTCAACGACCCCTCCGATCCGATCTTGACTCGCACTTGATGATTTCA 1252  
Db 1726 AGGTGCTCAACGAGTTCGCGCTGCACTACCGGACTATATCGCGCGATGTGG 1785  
Qy 1253 AGGCGATTACCGACGCGCGCTCGTCAAGGGGATCTTGCCTGGCGCTTGCCTGATA 1312  
Db 1786 ATGCTGTGACACGAGATGGGGTCAAGCTTACATGGCTTGGAGTTGTTAGAT 1845  
Qy 1313 ACTTGA 1319  
Db 1846 AAGTGA 1852

RESULT 11  
LOCUS AX654334 1404 bp DNA linear PAT 22-MAR-2003  
DEFINITION Sequence 4204 from Patent WO03000898.  
ACCESSION AX654334  
VERSION AX654334.1 GI:29157148  
KEYWORDS  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE 1  
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,  
Katsagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.  
TITLE Plant genes involved in defense against pathogens  
JOURNAL Patent: WO 03000898-A 4204 03-JAN-2003;  
Syngenta Participations AG (CH)  
FEATURES  
SOURCE 1..1404  
Location/Qualifiers  
/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

ORIGIN  
Query Match 16.9%; Score 246; DB 6; Length 1404;  
Best Local Similarity 54.5%; Pred. No. 4.6e-36;  
Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;

Db 497 ACAAGGTAAGGACTGGATACCGCTCAAGAGCCGACACGATGGCCATCCAGGGCTACG 556  
Qy 542 CCAACCGGACCAACGCGCCCGGAGAGAGCATTA-----ACAGACACTCCACGAGG 595  
Db 557 ACCGAGGGCTCCAGGGCCCGCGCGCTGCTCGTGTCTCCACTCTACTATGCAAGCCG 616  
Qy 596 GCAACACTGCACTGAGCCGTGGCTCGTGAAGAGCCGAGATCATAGGCTAGCCCGCG 655  
Db 617 GCAACTCCGACACGAGCCCTAGCTGTGCTGCCCAACACTTCACTCTCGCCACCGCCG 676  
Qy 656 CCGTGGCCGTCTACAGAGGAGCTTTCGCCCTCGCAAAAGGCGAGATCGCATCTCGC 715  
Db 677 CCGCCAGCATCTACAGGACAAATAACAGGCGACGAGAAAGGCGAGCTGGATAGCGT 726  
Qy 716 TCAACGGCACTACTATAGGCGCTGGAGACGAAATAGCCTCGGACAAAGAGGCTGCTG 775  
Db 737 TCGACGTGATGTGTTCGAGCCGATGTCCAAACCAAGAT--CGACATCGAGGCGGCA 793  
Qy 776 AGGACGATGAAATTTACATTTGAGCTGGTTTCCATCCCATCTTCTTGAAGAGACT 835  
Db 794 AGAGAGCGCAGAGCTTTCAGCTAGATGTGTCTGATCCGTTCTTT---CGGGAGCT 850  
Qy 836 ATCCAGAGACATGAAGAGACAGCTGCGGAGGCTTCCAGGCTTCCAGCTCCCGAGCT 895  
Db 851 ACCCGGACATGAGAGGCGAGGKTGGGGAGAGGCTGCCAGATTCAAGCGGATGAGG 910  
Qy 896 TTGCGATCTCATATGCCGAGAGACCGCATTTCTACGGCATGAATTTACTACATCCCACT 955  
Db 911 CCGCCGTGTCAAG---GGGGCCCTGGATTTCTGTGGCATMAACACTACACCACTACT 967  
Qy 956 TCGCGGCGCACTAGACGCTCCGTCGCCGAGAG-----CGACTATCTCGGCGCATCCA 1010  
Db 968 ACACGAGGCAACAACAACCAACATATCGGACATTTGCTCAACACACCTTGGACACA 1027  
Qy 1011 TGAACACCAAGAGAAAT-----AAGAGCGGACGCCCTGTGCGAGAGAGCGGCTCG 1063  
Db 1028 CCGGACCGCTCACTCCCATTCATCAAGAAATGGAAAGCAATTTGAGATAGGGCAATTCGA 1087  
Qy 1064 CTTGCTGCGCTCTCTCCCGGACATGTTCCGGAAGCATCTCGCCGGGTATAGGCTCT 1123  
Db 1088 TATGCTGTATCATTTGCCCCGAGGATGAGAGCTGTGAATGTATGCAAGAAAGT 1147  
Qy 1124 ---ACGGCAAGCCCATCTATCATCACCGAGAACGATGCCGTCCTCGAGAGAGACA 1180  
Db 1148 ACAACAGCCCAACAGGTATCATCTGAACAAACGGATGATGACAGACACACCTTTCA 1207  
Qy 1181 TGAAGTGCAGAGAGCGCTCAAGACCCCTTCGCAATCCGATCTTGAATCGCACTTGG 1240  
Db 1208 TTTTCATCAAGAGACGCTCAAGAGACAGCAAGAGATCAAAATACCAATGATCTTCA 1267  
Qy 1241 ACTCGATTTCAAGGCGATTAACCAAGAGCGGCTGTCTGCAAGGGGATCTTGGCTGG 1300  
Db 1268 CCAATCTGCTGCTTTCATTCMAAGAGACGGGTGCACAGTACGTGGGTATTTGGCGTGT 1327  
Qy 1301 CGTTGCTCATTAATTGAA--TGGTCAGATGGCTACGAGACCCGATTTGGCGCTACGTT 1359  
Db 1328 CTCTGCTGCAACTGGAAATGGCGGCGGATATCTCTCGAGATTCGGGCTCTACTTC 1387  
Qy 1360 ACAGACTTCA 1369  
Db 1388 GTGAGCTACA 1397

RESULT 12  
AK067231 2148 bp mRNA linear PLN 24-JUL-2003  
LOCUS AK067231  
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013095012, full  
insert sequence.  
ACCESSION AK067231  
VERSION AK067231.1 GI:32977249  
KEYWORDS FLI cDNA; CAP trapper.  
SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
Eriocaridaceae; Oryzoideae; Oryza.

REFERENCE

AUTHORS

The Rice Full-Length cDNA Consortium, National Institute of  
Agricultural Sciences Rice Full-Length cDNA Project Team;  
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,  
Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,  
Kojima, K., Namiki, T., Ohneda, E., Yanagi, W., Suzuki, K., Li, C.,  
Ohtsuka, K., Shishiki, T., Foundation of Advancement of International  
Science Sequencing & Analysis Group; Ootomo, Y., Murakami, K.,  
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,  
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,  
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikura, J.,  
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mita, J.,  
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;  
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,  
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,  
Kagawa, I., Kondo, S., Komu, H., Miyazaki, A., Otsu, N., Ota, Y.,  
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Yoshino, M., and Hayashizaki, Y.  
Collection, mapping, and annotation of over 28,000 cDNA clones from  
japonica rice  
Science 301 (5631), 376-379 (2003)

TITLE

2 (bases 1 to 2148)

JOURNAL

Medline  
22752273  
12699764

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of  
Agricultural Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp.  
Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica  
rice.  
URL: http://cdna01.dna.affrc.go.jp/cDNA/  
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J.,  
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, E., Yanagi, W., Suzuki, K., Li, C., Ohtsuka, K., Shishiki, T., and  
Yamamoto, M.  
FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Mita, J.,  
Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S.,  
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
Yoshimura, A., Matsubara, K., and Murakami, K.  
Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,

FEATURES

source

Hiraoaka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
Kishikawa, H., Kishimoto, T., Kojima, Y., Kondo, S., Komu, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohtsuka, N.,  
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sugabe, Y., Tagami, M., Tagami, T., Tagami, T., Tanaka, F.,  
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
Yasunishi, A., and Hayashizaki, Y.  
Location/Qualifiers  
1..2148  
1..2148  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J013095012"

ORIGIN

Query Match 15.2%; Score 221.6; DB 8; Length 2148;  
Best Local Similarity 55.9%; Pred. No. 1.6e-31;  
Matches 529; Conservative 0; Mismatches 399; Indels 18; Gaps 5;

20 TGCCCAACGACTTGAATGGGCTTCGCAACGCGCCCTACAGATCGAAGCGCGTCA 79  
203 TCCCGAGAGACTTATCTTCGCGACCGGCTCCGCGCTTATCGTACGAGGCGCTGTCA 262  
80 AAGAAGTGGCGCGCGCGCGCTTCATCTGAGACAGTACTCCACTGAGCCATCGCGCA 139  
263 ACAAGGCGCGCAAGGCGCGCGAGCATCTGGGACACGTACCTCACTTCCAGGCAAGTTG 322  
140 CCAAGCGCGCGCAAGGCGCGAGTGGCTTGGATACACCGCTACGATGAGACTTTG 199  
323 AAGATGTACCAAGGAGATGACGTGACTTTTACATGCTTACAGAGAGATCTGA 382  
200 ATCTTTGACCAAGTACGCGCGCAAGGCGCTTACGCTTCTTCTGCTGAGCGGATCA 259  
383 ACTTCGACCGACATGAAACATGAGACGCTTCGCTTCCATGCTTGGAGAGATCC 442  
260 TTCCCTCGCGCGCGAGCTTGATCCGCTACGAGAGAGAAATGATTTACAGCAAC 319  
443 TGCCAATGAAACATCACTGAGAGAAATCAAAAGAGGATTTCTTCTTCAACAGCC 502  
320 TGAATGACGCGCTTGAAGCGCGGATATGATGAGCTTGTGATCTTCACTGGGATC 379  
503 TGAATGACGCGCTTGAAGCGGATATGATGAGCTTGTGATCTTCACTGGGATC 562  
380 TGCCTGAGCGCTTCACTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATG 439  
563 CCCACAGGCTTGGAGAGCAAAATACCGGAGCTTCTCA--GTGAAACATTTGAAAG 619  
440 ACTTTGAGCGGATGAGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTT 499  
620 ATTTGATGATGAGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTG 679  
500 TCACATCAACGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATG 559  
680 AACAGTCAACGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATG 739  
560 CGGCGAGAGGCG 613  
740 CGGCGCGCGCTTGTGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGT 799  
614 CGGCG 673  
800 CGGCG 859  
674 GCGAGCTTGTGCG 733  
860 AAAAGTCAACGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATG 919  
734 AGCCCTGAGAGCGATGAGCTTGTGATGAGCTTGTGATGAGCTTGTGATGAGCTTGT 793

Db	920	TGCGCT---ACAGGAGACGCCGCCGCCGACAGCAGCAGCGCTGAGGCGACCTCGACTTCA	976
Qy	794	ACATTGCTGTTTGGCCAAATCCATCTTTCTTGAAGAGAACTATTCAGAGACATGAGA	853
Db	977	TGTACGGATGCTTATGATGACCCGATCTGTGTT---CGGCGACTTACCCGGGACATGTAGAGA	1033
Qy	854	AGCAGCTGGGGAGAGAGCTTCACGCCCTCATCTCCCGCGAGCTTTGGCATCTCTCAATGCC	913
Db	1034	AGTTGTGCGAGACCGCGCTGCCAAATTACAGCGCAGAGCTGCGAGTTGGTGA---CG	1090
Qy	914	GAGAGACCGACTTCTACGCGCATGAAATTACTACATATCCCACTTGGC	959
Db	1091	GATCTTGACATTTATGCGCTCAACATCTACACCACTTACGC	1136
RESULT 13			
AK120430		2148 bp	mrna
LOCUS			linear
DEFINITION	Oryza sativa (japonica cultivar-group)	CDNA clone:J013095M12, full	
ACCESSION	AK120430		
VERSION	AK120430.1	GI:37990053	
KEYWORDS	FLI CDNA; CAP trapper.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
REFERENCE			
1			
1	The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team, Kikuchi,S., Satoh,K., Nagata,T., Kawagashita,N., Doi,K., Kishimoto,N., Yazaki,Y., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Oono,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tanoda,Y., Kuroseki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., Marikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nilkura,J., Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Mura,Y., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN: Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arikawa,T., Fukuda,S., Hara,A., Haashizume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Oato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shitaki,T., Yoshino,M., and Hayashizaki,Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice		
JOURNAL	Science	301 (5631)	376-379 (2003)
MEDLINE	22752273		
PUBMED	12869764		
REFERENCE			
AUTHORS	2		
	Aachhi,J., Aizawa,K., Akimura,T., Arikawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanegaki,T., Hara,A., Haashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hizomoto,K., Hirooka,T., Horii,F., Hotta,I., Iida,J., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kawamata,S., Katoh,H., Kawagashita,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kuroseki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsumura,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakahara,Y., Nakamura,M., Namiki,T., Narikawa,R., Nilkura,J., Nishii,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Ooto,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shitaki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Tekeda,Y., Tagawa,A., Takahashi,F., Takai-Akita,S., Tanaka,T., Tomaru,A., Toyota,T., Tanoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,M., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S., and Yoshimura,A.		

JOURNAL	Collection, mapping, and annotation of 28K full-length cDNA clones from japonica rice
REFERENCE	3 (bases 1 to 2148)
AUTHORS	Kikuchi, S.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skk@kuchonias.affrc.go.jp, Tel:+81-29-838-7007, Fax:+81-29-838-7007)
COMMENT	This clone is one of the 32K full-length cDNA clones from japonica rice. URL : http://cdna01.dna.affrc.go.jp/cdna/NINS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Onitsuki,K., Shishiki,T., Yamamoto,M. and Nakahama,Y. FAS Genome Sequencing & Analysis Group: Otsuo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusunegi,T., Lu,M., Masuda,H., Mura,J., Mizuno,K., Narikawa,R., Nikiya,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tannoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K. Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akamura,T., Arakawa,T., Carninci,P., Fukuda,S., Hangaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hitoe,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itchihara,K., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Koude,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazeki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saichou,H., Sakai,C., Sakai,K., Sekazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Tanaka-Akahira,S., Tanaka,T., Tomaru,A., Toyo,T., waki,K., Yasunishi,A. and Hayashizaki,Y. Location/Qualifiers 1..2148 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /cultiyar="Nippondare" /db_xref="taxon:39947" /clone="J01309SM12"
FEATURES	
SOURCE	
ORIGIN	
Query Match	15.2%; Score 221.6; DB 8; Length 2148;
Best Local Similarity	55.9%; Pred. No. 1.6e-31;
Matches 529,	Conservative 0; Mismatches 399; Indels 18; Gaps 5,
Oy	20 TGCCCAACGACTTTGAATGGGCGCTTTCGCACAGCGCGCCTACCAGATCGAAGCGCGCTCA 79
Db	203 TCCCGGAGACTTCATCTTGCGCACCGGCTCCGGCGCTTAACAATGACAAAGCGCGGTGGA 262
Oy	80 AAGAAGTGGCGCGGCCCGGTCATCTTGCGGACAGTACTGCGACCTGGAGCATCGCGCA 139
Db	263 ACGAAGGCGGACAGGGCGCCAGCATCTGGGACACGTACGCTCACTTCACAGCAAATTG 322
Oy	140 CCAACGGCGCCAACGGCGGATGGCTTGGCATCTACTACACCGGCTAGCGATGAGGACTTTG 199
Db	323 AAGATGTGACCAAGGGGATGTACAGTAGACTTTTACCATCGCTCAACAAGAGGATGTGA 382
Oy	200 ATCTCTTGACCAAGTACGGCGCAAAAGGCTTACCGCTTCTCTTGTGCTGTGCGGATCA 259
Db	383 ACTTCGTGACCGAATGAACATGGACCGCTTTCCTTCTCATTTGTTGGAGCAGGATCC 442
Oy	260 TTCCCTTCGGCGGAGCGTGGATCCGTCACAAGGAGGAGGAATTGAGTTTACAGCAAC 319
Db	443 TGCCAATGGAACATCAGTGGAGGAATCAACAAAGGGGATTTGCTTTTCAACAAGCC 502

OY	320	TGATTGACGGCCTGTTGAGCGGGGATATACGCGCTTGAGTCTTGTACCATCTGGGATC	379
Db	503	TGATTACGAGGTCATATCTAGAGGGGTGAAGCCATTGTGCACCATCTTCATTTCACACA	562
OY	380	TGCTCAGGGGCTTACAGATCGCATATGAGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGG	439
Db	563	CCCACAGGCTCTTGAGAGCAAAATCCGAGCTTCTCA--GTGAAAACATTGTGAAGG	619
OY	440	ACTTTGAGCGGTATCGAGGTTGTGCTTTGAAACGTTTGGGAGCCGAGTCCAGAACTGA	499
Db	620	ATTTCGTGAGACTATGCGGAGCGTGTGCTTCGTGAGTTCCGGGACCGGGTGAAGTCATGA	679
OY	500	TCACCATCAGACGCCCTGATTAAGGCGCATCTATGATATGCAACCCGCGACGACGCC	559
Db	680	ACACGTTCAACGAGCCGATGATCTTCTGCGCGCGGCTACGAGCTCCGGGACCAAGGCC	739
OY	560	CGGGGACGAGG-----AGCATTAACAAGCACTCCACGAGGGCAACACTGCCACTAGC	613
Db	740	CCGGCCGCTGCTCCGCGTACGTTCTCAAGAAATCGCCCCCGGCGATCTCCGCAACGAGC	799
OY	614	CGTGGCTCTGTAAGAAAGGCCGATCATGATGAGCCATGCGCGCGCTGGCGCTTACAGCA	673
Db	800	CGTACGTGCGCGGCAACCTGCTGCTCGCCACGCGGAGCGCTCGCTTACCGGCC	859
OY	674	GGGACTTTCGCGCTCGGAAAAAGGCCGATCGGCGATCTCGCTCAACGGCGCATACTATG	733
Db	860	AAAAGTACCAAGCCAGCAGGAAGGGGAGATGCGATCAGCAGAGTGTGCGACTGTTCC	919
OY	734	AGCCCTGGGACAGCATGAGCTCGGAGCAAGAGGCTGTGAGAGCGATGGAATTTTC	793
Db	920	TGCGCT---ACAGCAGCGCCCGCCCGGACAGACGCGGTGAGGGCGAGCTTCGACTTCA	976
OY	794	ACATTGCGTGTGTTGCCAATCCCATCTTCTTGAGAGAGCATATTCAGAGGCGATGAGA	853
Db	977	TGTACGAGATGTTCTATGACCCGATCGTGT---CGGCGACTTACCCGCGCACCATGAGA	1033
OY	854	AGCAGCTGGGAGAGGCTTCCAGCCGTCACCTCCGCGGAGCTTTCGATCTCCATGTCGC	913
Db	1034	AGTTGTGGGTACCGGCTGCCCAATTACAGCCGAGCATGCGAGTTGGTGA---GG	1090
OY	914	GAGAGACCGACTTCTACGCGATGAATTAATCACTCCCACTTCGC	959
Db	1091	GATCTAGACCTTCACTCGGCTCACTCACTACACCAACTACGC	1136
RESULT 14			
AK066710			
LOCUS			
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:J013074A20, full insert sequence.	1811 bp	mRNA linear PLN 24-JUL-2003
ACCESSION	AK066710		
VERSION	AK066710.1	GI:32976728	
KEYWORDS	FLI CDNA; CAP trapper.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.		
REFERENCE			
AUTHORS	1		
	The Rice Full-Length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team; Kikuchi S., Satoh K., Nagata R., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Ootomo Y., Murakami K., Iida Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y., Kuroseki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M., Narikawa R., Sugiyama A., Mizuno K., Yokomizo S., Nikura J., Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J., Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN, Kawai J., Carninci P., Adachi U., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashidume W., Hayatsu N., Imotani K., Ishii Y., Itoh M.,		
TITLE	Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice Science 301 (5631), 376-379 (2003)		
JOURNAL	MEDLINE		
PUBMED	22752273		
REFERENCE	12869764		
AUTHORS	2 (bases 1 to 1811)		
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kuroseki, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kuroseki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, R., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)		
COMMENT	This clone is one of the 28K full-length cDNA clones from japonica rice.		
	URL: http://cdna01.dna.affrc.go.jp/cDNA/NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and Yamamoto, M.		
	FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kuroseki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K., and Murakami, K.		
	Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, R., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A., and Hayashizaki, Y.		
FEATURES	Location/Qualifiers		
source	1..1811		
	/organism="Oryza sativa (japonica cultivar-group)"		
	/mol_type="mRNA"		
	/cultiivar="Nipponbare"		
	/db_xref="taxon:39947"		
	/clone="J013074A20"		
ORIGIN			

Query Match 14.8%; Score 215.8; DB 8; Length 1811;  
 Best Local Similarity 51.3%; Pred. No. 2e-30;  
 Matches 714; Conservative 1; Mismatches 648; Indels 30; Gaps 8;

QY 20 TGGCCACGACCTTGAATGGGGCTTGGCAACGGCCGCTTACAGATGGAAGGGCCCTCA 79  
 DB 135 TCCCAAGGACTTCATCTTCGGGCAACAGTTTCAGCAGCTTATCAGTATAGAGGGCCCTTACA 194  
 QY 80 AAGAAGTGGCCGGCGCCGCTTCATGGGACACGTACGCAACCTGAGCCATTCGGCA 139  
 DB 195 AAGAAGGGGGCAAGGCTTCAAGGCTTGGGACAACTTCACTCACTCAAGTAAATTT 254  
 QY 140 CCAAGCGCCCAACGGCGATGTGGCTTGGCATCTACCAACCGCTAGATAGAGACTTTG 199  
 DB 255 TAAACAATGATTAACGGGATGTGGCAATGACTTCTATCAACGATACAAAGAGATGTGA 314  
 QY 200 ATCTCTTGAACAAAGTACGGGCAAGGCTTACGGCTTCTCTGTGTGGTGGATCA 259  
 DB 315 GCTCTCTGAAGACATGAAATGATGATCTTCCGGTCTCCATTTGCTGGAACAGATCC 374  
 QY 260 TTCCCTCGGGGAGGCTGGATCCGTCACAGAGAGGAATTGAGTTTACAGCAAC 319  
 DB 375 TCCCAATGATCTTGTAGTGAAGAAATAACAAAGAGGGTGTCTTTTACAAACGCT 434  
 QY 320 TGAATGACGCTTGTGAGGCGGGGTATCAAGCCTTGGGTGACTTTGTACACTGGATC 379  
 DB 435 TGATCAATGATGTATGCAAAAGGATGATCCATTTGTCACTATCTTCACTGGGACA 494  
 QY 380 TGCCTCAGCGCTTACAGATCGCTATGAGAGGCTGCTCAACGTGAAAGGTCAGCTGG 439  
 DB 495 CCCCCTTGGCTTGAAGAAAGCAATACGAGGATTTCTCA---GTGAAGACATAGTGAAG 551  
 QY 440 ACTTTGAGCGGTATGCGAGGTGTGCTTGAACCTTTTGGGACCGGACTCAAGACTGGA 499  
 DB 552 AATACGTGACTTTCGCGAGGTGTCTTTCGCGATTCGGGACCGGTCAAGTACGATGGA 611  
 QY 500 TCAACATCAACGACCCCTGATTCAGGCACTATGATATGCAACCGGAGCAACGCC 559  
 DB 612 CCACATTTAATGAGCCATTCAATATAGCCCTTACGCTACGCAAGGGCGTGTTCGCG 671  
 QY 560 CGGGACGAGACGACA-----TTAACAAGCACTTCACCGAGGCAACACTGCGCATGAGC 613  
 DB 672 CGGAGACGATCTCTTCATATGTTTCCAAATCATGCGCGCTGCGGATCTCAAGTGCAGC 731  
 QY 614 CGTGGCTCGCTGGAAGGCCCAATCATGAGCCATGCCCCGCGTGGCCGTCTACAGCA 673  
 DB 732 CTAACCTCTGGGACACACATCCACTCTCCACGCTGCGCTGTCCAGCTTTTACCGCA 791  
 QY 674 GGAAGCTTTCGCCCCCTCGCAAAAGGGCCAGATCGGATCTGCTCAACGGGCACTACTATG 733  
 DB 792 CCAAGTACAGCCCAACACAGAGGACAGATCGGATGTGTGTGTACCACTGCTTTCG 851  
 QY 734 AGCCTTGGGACAGCAATGAGCTTCGGGACAAGAGGCTGTGAGCAAGGATGGAATTC 793  
 DB 852 TGGCGTACGACCACTCGACGCT---GACCGTGGCGCTGTGCAACGAGCCTTAGACTTCA 908  
 QY 794 ACATTTGGCTGTGGTCCCAATCCCATCTTCTTGAAGAAAGACTATCCAGAGACTTGAAGA 853  
 DB 909 TCTATGGGTGTTCATGAGACCTTA--TGTGTGATGTGTGACTTACCCAGGACCAATGAG 965  
 QY 854 AGAGAGTGGGAGAGGCTTTCAGACCTCACTCCCGGGAATTGGCATCTCAATCCG 913  
 DB 966 GTTGGCTCGGCATTCGGCTGCGAGATTCACGCTTGAACAGTGGCCATGTGTAAAGGCT 1025  
 QY 914 GAGAGACGACTTCTACGGAATGAAATTAACATCCAGTTTCGCGGCAACCTTAGACG 973  
 DB 1026 CCTA---GCACTTCATCGGATTAATTAACACCACTACTAGCTTAAGATTAACGCG 1082  
 QY 974 GTCCCGTCCCGAGAGCG-----ACTATCTCGCGGCACTCATGAGACCAAGAGATA 1027  
 DB 1083 CGCTTAATCTCAACGAGACTTCTTACGACCTTGAACAACCGGCAACACACCGGCTTCC 1142  
 QY 1028 AGGACGGACAGCCCGTGGGAGAGAGGCGCTCGCTGGCTGCGCTCTGCGCGGACA 1087

DB 1143 GTAATGGCAAAACCCATCGGTCCACAGGAATTTACACCGATCTTCTCACTACCTTCAG 1202  
 QY 1088 TGTTCGGGAAGCATTCGCCCCGGGTGTACGGCCTGTACAGCCAGCCCATTAATACCG 1147  
 DB 1203 GTCTCGGTGAGCTTCCTCTCTTACCAAGAGAGATTAACAACCCGA---CCATTTATG 1259  
 QY 1148 AAGAAGTATGCGCGTCCGTGAGAGAGAACATGACGTGCGAGAGGCGGTCAACGACC 1207  
 DB 1260 TTACAGAAAACGCGATCATGATGAGGTATACACAGCACACTGCCAGAGGCGCTCAAGATG 1319  
 QY 1208 CTTTCGCACTCCGTAATTTGACTCGCACTTGTGACTGTGATTTTCAAGGCCATTACCAG 1267  
 DB 1320 GACACAGATTCGATTCACATCAAGGACCTGCAAGTTTCGTCACACGCGCAT---CAAGA 1376  
 QY 1268 ACGCGCTCTGTCTCAAGGGGTACTTTCGCTGGCGCTTGTCTGTATTAATTGGAATGTGAC 1327  
 DB 1377 ATGGGGTGAACGTGAAGGGCTTACTTCACTGACATTCATGAGCTGTTGAGTGGGGTG 1436  
 QY 1328 ATGGCTACGGAACCCAGATTCGGCGTCAAGTTCAAGACTACACACCTCAAGCGCACG 1387  
 DB 1437 ACGGCTACCTTGAACAGGTTGCGCTCATCTACGTCAACGCAAGACGCTCAAGCGCTACC 1496  
 QY 1388 CCAAGAGTCTGC 1400  
 DB 1497 GCAAGAGTCAAGC 1509

RESULT 15  
 AX652977 1503 bp DNA linear PAT 22-MAR-2003  
 LOCUS AX652977  
 DEFINITION Sequence 2847 from Patent WO03000898.  
 ACCESSION AX652977  
 VERSION AX652977.1 GI:29155791  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa  
 Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE  
 AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,  
 Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.  
 TITLE Plant genes involved in defense against pathogens  
 JOURNAL Patent: WO 03000898-A 2847 03-JUN-2003;  
 FEATURES  
 location/Qualifiers  
 1..1503  
 /organism="Oryza sativa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:4530"

## ORIGIN

Query Match 14.7%; Score 214.6; DB 6; Length 1503;  
 Best Local Similarity 51.0%; Pred. No. 3.4e-30;  
 Matches 720; Conservative 1; Mismatches 660; Indels 30; Gaps 8;

QY 20 TGGCCACGACCTTGAATGGGGCTTGGCAACGGCCGCTTACAGATGGAAGGGCCCTCA 79  
 DB 95 TCCCAAGGACTTCATCTTCGGGACAGAGTTTCAGACCTTATCAGTATGAGGGGCGCTTACA 154  
 QY 80 AAGAAGTGGCGCGGCGCGCTTCATCTGGGACAGTATGCGACCTGAGAGCATTCGGCA 139  
 DB 155 AAGAAGGGGGCAAGGCTTCAAGGCTTGGGACAACTTCACTCACTTCCAGTAAATTT 214  
 QY 140 CCAAGCGCCCAACGGCGATGTGGCTTGGCATCTACCAACCGCTAGATGAGAGACTTTG 199  
 DB 215 TAAACAATGATTAACGGGATGTGGCAAAATGACTTCTATACCGATACAAAGAGAGATGA 274  
 QY 200 ATCTCTTGAACAAAGTACGGGCAAGGCTTACCGCTTCTCTTGTGTGTGCGGATCA 259  
 DB 275 GCTCTCTGAAGGACATGAACATGATGATCTTCCGGTCTTCATGTGCTGAGCAACGAGATCC 334

QY 260 TTCCCTCGCGGAGGCTGGATCCCGTCAACGAGAGGGAATTGAGTTTACAGCAAC 319  
 Db 335 TGCCAAATGATCTTATGAGTGAAGATTAACAAGAGAGGAGTTGCTTTTCAACAGCT 394  
 QY 320 TGATTTACGCGCTGTGAGGCGGGTATCAAGCTTGGGTGATCTTTGACCTATGGGATC 379  
 Db 395 TGATCAATGATGTATGCAAAAGGAGTATCCATTTGTCTATTTTCACTGGGACA 454  
 QY 380 TGCCTCAGGCGCTTCAAGATGCTATGAGGCTGGCTCAACGTGGAAGAGGTCCAGCTGG 439  
 Db 455 CCCCCTTGGCTCTGGAAGCAAAATACGAGGATTCCTCA--GTGAAGACATAGTGAAG 511  
 QY 440 ACTTTAGCGGTATGCGAGGTTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAACTGA 499  
 Db 512 AATAGGTGACTTGGCGAGGAGTGTGCTTCCGAGTTCCGAGACCGTGTCAAGTCTGA 571  
 QY 500 TCACATCAAGAACCCGATTCAGGCTATGATATGCAACCGGAGCAACGCCC 559  
 Db 572 CCACATTATGAGCATTTCATATAGCGGCTACGGCTACGGAAGGCGGTGTTGGCG 631  
 QY 560 CGGCGAGGAGCA--TTAACAACATCCACCGAGGCAACTGCTCAAGC 613  
 Db 632 CGGAGCATGCTCTTATATGTTTCAATGCTTCAATGCGGCTGCTGACTCCAGTCCGAGC 691  
 QY 614 CGTGCTGCTGGAAGGCGCAGATCATGAGCCATGCCGCGGTGGCGGTCTACAGCA 673  
 Db 692 CTTACCTGCTGACACACACATCCACTTCCACGCTGCGCTGTCCAGCTTTACCGCA 751  
 QY 674 GGGACTTTGCGCCCTCGCAAAAGGCGCAGATCGGATCTGCTCAACGCGCATATG 733  
 Db 752 CCAAGTACAGCAACACAGAAAGGAGCATGCGGATGAGTGTGCTACCCACTGTTGCG 811  
 QY 734 AGCCCTGGAGCAACATGAGCTCGGACAGAGAGCTGTGAGGACGAGATGGAATTTC 793  
 Db 812 TGCCGTACGACCACTCCGAGCT--GACCGTGGCGCTGTGCAACGAGCCTAGACTTCA 868  
 QY 794 ACATTTGGCTGTTTGGCAATCCCATCTTTTGAAGAAGGATATCCAGAGCATGAAGA 853  
 Db 869 TCTATGGTGTTCATGAGCCTA--TGTGCAATGTGATCAACCAAGCAACATGAGAG 925  
 QY 854 AGCAGCTGGGCGAGAGGCTTCCAGCCCTGACCTCCGCGGACTTTGGCATCTCAATGCCG 913  
 Db 926 GTTGGCTGGCAATCGGCTGCCAGAGTTCAGCCTGACAGTGGGATGAGTGAAGGCT 985  
 QY 914 GAGAGCCGACTTCTACGCGCATGAATTAATACACATCCAGTTCCGCGCGCACCTAGAGC 973  
 Db 986 CCTA---CGACTTCATGGAATTAATTAATTAACACACTTACCTAGATTAACCGC 1042  
 QY 974 GTCCCGTCCCGA-----GACGACTATCTCGGCGCATTCATGAGCACAGAGATA 1027  
 Db 1043 CGCTTAATCTCAACAGCTATCTTAAGACCTGAGCAACCGCGCAACACCGGCTTCC 1102  
 QY 1028 AGGACGCGAGCCCGTTGCGAGAGAGAGCGGCTCCGCTGGCTGCTGCGCGGACA 1087  
 Db 1103 GTATGCGCAAAACCATGCTGCCAGAGAAATTTACACGATCTTCTTAACCTACCTCCAG 1162  
 QY 1088 TGTTCGAGAGCATTCGCGCGGAGTGAAGGCTGTACGCGAGCCATCTACATCACCG 1147  
 Db 1163 GTCTCGTGAAGCTCTCTTCAACCAAGAGATTAACAAACCGA---CCATCTATG 1219  
 QY 1148 AGAAGGATCCCGTGCCTGAGAGAGAAATGACGTGCGAGAGGCGGTCAAGACC 1207  
 Db 1220 TTACAGAAAGAGGATGATGAGGTAAACAAGACACTGCGAGAGGCGCTCAAGAGATG 1279  
 QY 1208 CTTTCGCAATCCGTACTTTGACTGCGACTTGGACTGATTTCCAGGCGCATTAACCGAG 1267  
 Db 1280 GACACAGGATCGAGTTCACCTCAAGACCTGCGAGTGTCAACACGCGAT--CAAGA 1336  
 QY 1268 ACCGCGTGTGTCAGAGGAGTCTTTGCGTGGCGTGTGCTGATTAACCTTGAATGTGAG 1327  
 Db 1337 ATGGGAGTGAAGTGAAGGCTACTTCAAGTGAATTCATGAGCTGCTTTGAGTGGGCTG 1396  
 QY 1328 ATGGCTACGACCAAGATTGGCGCTCACTTCAACAGACTACCACTTCAAGGCGACGC 1387

Db 1397 AGGCTTACCTTGAACAGGTTCCGCGCTCATCTACGTGACCGCAAGAGCTCAAGGCTTACC 1456  
 QY 1388 CCAGAGTCTGCCCTGCTGCTCAAGACAT 1418  
 Db 1457 GCAGAGTCCAGCTACTGATGAGTGAAGACTT 1487

Search completed: March 28, 2004, 20:52:13  
 Job time : 5767.87 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 19:13:22 ; Search time 576.342 Seconds  
(without alignments)  
10724.755 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455  
Sequence: 1 atgcccagctgcctagctctc.....gggttaagtgcggcataa 1455

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapevt 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1454.2	99.9	1455	9	ADD22904 T. reesei
2	1440.6	99.0	1990	9	ADD22902 Full leng
3	598	41.1	735	3	Aaf14935 Trichoder
4	266.6	18.3	1933	3	Aaf14882 Trichoder
5	246	16.9	1404	7	ADA70881 Rice gene
6	214.6	14.7	1503	7	ADA69524 Rice gene
7	205.4	14.1	575	7	ABZ53130 Aspergill
8	182.8	12.6	1734	6	ABZ13477 Arabidops
9	182.8	12.6	1734	7	ADA67833 Arabidops
10	178.6	12.3	1533	4	ADA70381 Rice gene
11	178.2	12.2	2016	4	AAS21370 Human CDN
12	178.2	12.2	2016	6	ABL68245 Human PRO
13	178.2	12.2	2016	6	ABL95734 Human ang
14	178.2	12.2	2016	6	ACD23979 Human hum
15	178.2	12.2	2016	7	ACA67120 CDNA enco
16	178.2	12.2	2016	7	ACA03729 CDNA enco
17	178.2	12.2	2016	7	ABX89267 DNA enco
18	178.2	12.2	2016	7	ACD41921 Human sec
19	178.2	12.2	2016	7	ACA04150 Human CDN
20	178.2	12.2	2016	8	ADA45772 Novel hum
21	178.2	12.2	2016	8	ADA76203 Human PRO
22	178.2	12.2	2016	8	ADA18853 Human PRO
23	178.2	12.2	2016	8	ADA61476 Homo sapi

## ALIGNMENTS

24	178.2	12.2	2016	8	ADB19261	Adb19261 Novel hum
25	178.2	12.2	2016	8	ADB27802	Adb27802 CDNA enco
26	178.2	12.2	2016	8	ADA62821	Ada62821 Novel hum
27	178.2	12.2	2016	8	ADB15845	Adb15845 Human PRO
28	178.2	12.2	2016	8	ADA47631	Ada47631 Human PRO
29	178.2	12.2	2016	8	ADA67426	Ada67426 Human PRO
30	178.2	12.2	2016	8	ADB30433	Adb30433 CDNA enco
31	178.2	12.2	2016	8	ADA85729	Ada85729 Novel hum
32	178.2	12.2	2016	8	ADA96941	Ada96941 Human PRO
33	178.2	12.2	2016	8	ADA79245	Ada79245 Human PRO
34	178.2	12.2	2016	8	ADA87384	Ada87384 Novel hum
35	178.2	12.2	2016	8	ADB16586	Adb16586 Human PRO
36	178.2	12.2	2016	8	ADA91678	Ada91678 Novel hum
37	178.2	12.2	2016	8	ADB14741	Adb14741 Human PRO
38	178.2	12.2	2016	8	ADB18702	Adb18702 Novel hum
39	178.2	12.2	2016	8	ADA93917	Ada93917 Human PRO
40	178.2	12.2	2016	8	ADB19813	Adb19813 Novel hum
41	178.2	12.2	2016	8	ADB13125	Adb13125 Human PRO
42	178.2	12.2	2016	8	ACD98550	AcD98550 Novel hum
43	178.2	12.2	2016	8	ADA74379	Ada74379 Human PRO
44	178.2	12.2	2016	8	ADB24612	Adb24612 Human PRO
45	178.2	12.2	2016	8	ADA82136	Ada82136 Human PRO

RESULT 1	ADD22904	standard; CDNA; 1455 BP.
ID	ADD22904	
XX	XX	
AC	ADD22904;	
XX	XX	
DT	15-JUN-2004	(first entry)
XX	XX	
DE	T. reesei bgl5 beta-glucosidase CDNA.	
XX	XX	
KW	BGL5; beta-glucosidase; fungus; ethanol; sugar;	
KW	fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;	
KW	aroma; detergent; softening agent; cotton; wood pulp; gene; ss.	
XX	XX	
OS	Hypocrea jecorina.	
XX	XX	
FH	Key	Location/Qualifiers
FT	CDS	1..1455
FT		/tag= a
FT		/product= "BGL5 beta galactosidase"
FT		/transl_except= (pos:37..39,aa:Gln)
FT		/transl_except= (pos:511..513,aa:Xaa)
FT		/transl_except= (pos:1219..1221,aa:Pro)
FT		/note= "Xaa= unknown"
XX	XX	
PN	US2003114330-A1.	
XX	XX	
PD	19-JUN-2003.	
XX	XX	
PF	18-DEC-2001; 2001US-00026140.	
XX	XX	
PR	18-DEC-2001; 2001US-00026140.	
XX	XX	
PA	(DUNN/) DUNN-COLEMAN N.	
PA	(GOED/) GOEDEGEBOUR F.	
PA	(WARD/) WARD M.	
PA	(YAO/) YAO J.	
XX	XX	
PI	Dunn-Coleman N, Goedegebour F, Ward M, Yao J;	
XX	XX	
DR	WPI: 2003-810909/76.	
XX	XX	
DR	P-PDB; ADD22903.	
XX	XX	
PT	New substantially purified BGL5 polypeptide with the biological activity	
XX	XX	
PT	of beta-glucosidase, useful for degrading biomasses to ethanol.	

PS Claim 2; Page 16-17; 21bp; English.

XX The invention discloses a substantially purified BGL5 polypeptide with  
CC the biological activity of a beta-glucosidase. Also claimed is an  
CC isolated polynucleotide derived from a fungal source, which comprises a  
CC nucleotide sequence encoding an enzyme having beta-glucosidase activity,  
CC a Trichoderma reesei bgl5 nucleic acid coding sequence, or its  
CC complement, an expression construct including the T. reesei bgl5 cDNA  
CC sequence, a purified recombinant enzyme having beta-glucosidase activity,  
CC a recombinant host cell comprising a deletion or insertion or other  
CC alteration in the bgl5 gene which inactivates the gene and prevents BGL5  
CC polypeptide production, an antisense oligonucleotide, a detergent  
CC composition comprising the polypeptide and expressing a heterologous  
CC polypeptide having beta-glucosidase activity in an Aspergillus sp. BGL5  
CC is useful for producing ethanol, which involves contacting a biomass  
CC composition with an enzymatic composition comprising the polypeptide to  
CC yield a sugar solution, adding to the sugar solution a fermentative  
CC microorganism and culturing the fermentative microorganism under  
CC conditions sufficient to produce ethanol, where the biomass composition  
CC may be optionally pretreated. The method further involves the addition of  
CC at least one endoglucanase or cellobiohydrolase. The pretreatment is with  
CC a dilute acid. BGL5 is useful in wine making for enhancing the potential  
CC aroma of the finished wine product. The detergent is useful as softening  
CC agent and also useful for improving the feel of cotton fabrics and for  
CC degrading wood pulp into sugars. The sequence presented is the T. reesei  
CC bgl5 cDNA.

XX Sequence 1455 BP; 312 A; 444 C; 423 G; 274 T; 0 U; 2 Other;

Query Match 99.9%; Score 1454.2; DB 9; Length 1455;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCCAGTGTGAGCTCTGCTCCCAAGCACTTTGAATGGGGCTTCGCAACGGCGCCCTAC 60  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 CAGATCGAAGCGCGCGTCAAGAAAGTGGCGCGCGCGCGCTCATCTGGGACACGTACTGC 120  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 CAGATCGAAGCGCGCGTCAAGAAAGTGGCGCGCGCGCGCTCATCTGGGACACGTACTGC 120  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 CACCTGAGAGCATGCGGACCAACGCGCGCAACGCGCATGTGGCTTGGATCACTTACAC 180  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 CACCTGAGAGCATGCGGACCAACGCGCGCAACGCGCATGTGGCTTGGATCACTTACAC 180  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 CGTACGATGAGGACTTGTATCTCTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCC 240  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 CGTACGATGAGGACTTGTATCTCTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCC 240  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 TTGTGCTGAGTGGCGGATCATTCCTCGCGCGGAGGCTGATCCCGTCAAGAGAGGGA 300  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 TTGTGCTGAGTGGCGGATCATTCCTCGCGCGGAGGCTGATCCCGTCAAGAGAGGGA 300  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 ATTGAGTTTACAGCAACCTGATGAGCGCCCTGTGAGGCGGGGTATCAACGCTTGGGTG 360  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 301 ATTGAGTTTACAGCAACCTGATGAGCGCCCTGTGAGGCGGGGTATCAACGCTTGGGTG 360  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 ACTTTGATCACTGAGGATCTGCTCAAGGCGTTACAGATGCTATGAGAGGCTGCTCAAC 420  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 361 ACTTTGATCACTGAGGATCTGCTCAAGGCGTTACAGATGCTATGAGAGGCTGCTCAAC 420  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 GTGAGAGAGTCCAGCTGAGCTTTGAGCGGATGCGAGGTTGTCTTTGAACGTTTTGGG 480  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 421 GTGAGAGAGTCCAGCTGAGCTTTGAGCGGATGCGAGGTTGTCTTTGAACGTTTTGGG 480  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 481 GACCGAGTCCAGAACTGATCAACATCAACGACCCCTGATTCAGGCCATATATATAT 540  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 481 GACCGAGTCCAGAACTGATCAACATCAACGACCCCTGATTCAGGCCATATATATAT 540  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 541 GCGACCGGAGCAACGCGCGGAGGAGGAGCACTTAAACAAGCACTCCACGAGGGCAAC 600  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 541 GCGACCGGAGCAACGCGCGGAGGAGGAGCACTTAAACAAGCACTCCACGAGGGCAAC 600  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCAGATCATGAGCCATGCCCGCGGTG 660  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 601 ACTGCCACTGAGCGGTGCTCGCTGGAAGAGCCAGATCATGAGCCATGCCCGCGGTG 660  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 661 GCCGTTCACAGCAGGAGCTTTGCGCCCTCGCAAAAGGCGCAATGCGCATCTGCTCAAC 720  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 661 GCCGTTCACAGCAGGAGCTTTGCGCCCTCGCAAAAGGCGCAATGCGCATCTGCTCAAC 720  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 721 GCGGACTACTATAGCCCTGAGGACAGCAATGACCTTCGGGACAAAGAGGCTGCGAGCGA 780  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 721 GCGGACTACTATAGCCCTGAGGACAGCAATGACCTTCGGGACAAAGAGGCTGCGAGCGA 780  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 781 CCGATGAAATTTACATTTGCTGTTTGCATATCCATCTTCTTGAAGAGACTATCCA 840  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 781 CCGATGAAATTTACATTTGCTGTTTGCATATCCATCTTCTTGAAGAGACTATCCA 840  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 841 GAGAGCATGAAGAGCACTGGGCGAGAGGCTTCAGCCTCACTCCCGGACTTTGCC 900  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 841 GAGAGCATGAAGAGCACTGGGCGAGAGGCTTCAGCCTCACTCCCGGACTTTGCC 900  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 901 ATCCCTCAATGCGGAGAGACCGAATCTTACGAGATGAATTAACAATCCAGTTCCG 960  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 901 ATCCCTCAATGCGGAGAGACCGAATCTTACGAGATGAATTAACAATCCAGTTCCG 960  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 961 CGCCACCTAGACGGTCCCGTCCCGGAGACGGAATCTCGGCGCCCATCCATGAGCACAG 1020  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 961 CGCCACCTAGACGGTCCCGTCCCGGAGACGGAATCTCGGCGCCCATCCATGAGCACAG 1020  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1021 GAGATTAAGAGCGGACGCGCCCTGTTGGCGAGAGAGCGGCTCGCTGCTGCTCTGC 1080  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1021 GAGATTAAGAGCGGACGCGCCCTGTTGGCGAGAGAGCGGCTCGCTGCTGCTCTGC 1080  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1081 CCGGACATGTTCCGGAAGACATCTGCGCGGCTGTACGAGGCTGTACGCGCAAGCCATCTAC 1140  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1081 CCGGACATGTTCCGGAAGACATCTGCGCGGCTGTACGAGGCTGTACGCGCAAGCCATCTAC 1140  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1141 ATCAGCGAAGAGGATGCGCGTCCCTGTGAGAGAGAACATGACGTGCGAGAGCGCGTC 1200  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1141 ATCAGCGAAGAGGATGCGCGTCCCTGTGAGAGAGAACATGACGTGCGAGAGCGCGTC 1200  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1201 AACGACCCCTTCCGATCCGATCTTGTACTGCGCATTTGAGATTCGATTTTCAAGGCCATT 1260  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1201 AACGACCCCTTCCGATCCGATCTTGTACTGCGCATTTGAGATTCGATTTTCAAGGCCATT 1260  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1261 ACCGAGAGCGGCGTGGTGTCAAGGGGTACTTTGCGGCGGCTTGCATTAATTGGAA 1320  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1261 ACCGAGAGCGGCGTGGTGTCAAGGGGTACTTTGCGGCGGCTTGCATTAATTGGAA 1320  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1321 TTGTCAGATGCTTACGGAACCCAGATTCGCGGTACGTTCAAGACTTACACACCTCAAG 1380  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1321 TTGTCAGATGCTTACGGAACCCAGATTCGCGGTACGTTCAAGACTTACACACCTCAAG 1380  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1381 CGCAGCGCCCAAGAGTGTGCTGCTGCTTCAAGAGCATGTTTCCGCGCGGAGAGGTT 1440  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1381 CGCAGCGCCCAAGAGTGTGCTGCTGCTTCAAGAGCATGTTTCCGCGCGGAGAGGTT 1440  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1441 AAAGTGGCGGATTA 1455  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 1441 AAAGTGGCGGATTA 1455  
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2

ADD22902 standard; cDNA; 1990 BP.

ADD22902;

15-JAN-2004 (first entry)

Full length T. reesei bgl5 beta-glucosidase cDNA.

BGL5; beta-glucosidase; fungus; ethanol; sugar;  
fermentative microorganism; endoglucanase; cellobiohydrolase; wine making;



Db 62 ATGCCCGAGTCGCTAGCTCTGCCCAAGACTTTGAATGGGCTTCGCAACGGCCGCTAC 121

OY	6	AGATTCGAAGCCGCGTCAAGAAAGGTGGCGCGGCCCGTCTCACTGTGGACAACGTACTGCG	120
Db	122	CAGATTCGAAGCCGCGTCAAGAAAGGTGGCGCGGCCCGTCTCACTGTGGACAACGTACTGCG	181
OY	121	CACCTGGAGCCATTCGCGACCAACCGCGCCACCGCGCATGTGGCTTGCATCACTACAC	180
Db	182	CACCTGGAGCCATTCGCGACCAACCGCGCCACCGCGCATGTGGCTTGCATCACTACAC	241
OY	181	CGCTACGATGAGACTTTGATCTCTTTACCAAGTACGGCGCAAGGGCTTACCGCTTCTCC	240
Db	242	CGCTACGATGAGACTTTGATCTCTTTACCAAGTACGGCGCAAGGGCTTACCGCTTCTCC	301
OY	241	TTTGGCGGTGCGGGATCATTTCCCTGGGGGGACGGGTGGATCCCGTCAACGAGAGGGGA	300
Db	302	TTTGGCGGTGCGGGATCATTTCCCTGGGGGGACGGGTGGATCCCGTCAACGAGAGGGGA	361
OY	301	ATTGAGTTTAAAGCAACTGATTTGACGCCCTGTGGAGCGGGGTATCAGCCTTTGGGTG	360
Db	362	ATTGAGTTTAAAGCAACTGATTTGACGCCCTGTGGAGCGGGGTATCAGCCTTTGGGTG	421
OY	361	ACTTTGTACCACTGGGATCTTGCTCAGCGCTTACGATGGCTATGAGAGCTGGCTCAAC	420
Db	422	ACTTTGTACCACTGGGATCTTGCTCAGCGCTTACGATGGCTATGAGAGCTGGCTCAAC	481
OY	421	GTGGAAGAAGTCCAGCTGGACTTTGAGCGGTATGCGAGGTGGCTTTGACGTTTTGGG	480
Db	482	GTGGAAGAAGTCCAGCTGGACTTTGAGCGGTATGCGAGGTGGCTTTGACGTTTTGGG	541
OY	481	GACCGAGTCCAGACTGTGATATCACCATAACGAMCCCTGGAATTACGCCATCTATGATAT	540
Db	542	GACCGAGTCCAGACTGTGATATCACCATAACGAMCCCTGGAATTACGCCATCTATGATAT	601
OY	541	GCCACCGGCGACCAACGCCCGGGGAGAGAGCATTTAAACGACTCCACCGAGGGCAAC	600
Db	602	GCCACCGGCGACCAACGCCCGGGGAGAGAGCATTTAAACGACTCCACCGAGGGCAAC	661
OY	601	ACTGCCACTGAGCCGTGGCTTCGCTGTAAGAAAGGCCCAAGTACAGAACCATGCCGGCGGTG	660
Db	662	ACTGCCACTGAGCCGTGGCTTCGCTGTAAGAAAGGCCCAAGTACAGAACCATGCCGGCGGTG	721
OY	661	GCCGTCTACAGCAGAGACTTTCCGCCCTCGCAAAAGGGCCAGATCGGCATCTCGCTCAAC	720
Db	722	GCCGTCTACAGCAGAGACTTTCCGCCCTCGCAAAAGGGCCAGATCGGCATCTCGCTCAAC	781
OY	721	GGCGACTACTATGAGCCCTGGGACAGCAATGAGCCTCGGACCAAGAGAGCTGTGAGCGGA	780
Db	782	GGCGACTACTATGAGCCCTGGGACAGCAATGAGCCTCGGACCAAGAGAGCTGTGAGCGGA	841
OY	781	CGGATGGAATTTCACTATTGGCTGGTTTGGCCATCTCCCATCTTTCTTGAAGAGAAGACTATCA	840
Db	842	CGGATGGAATTTCACTATTGGCTGGTTTGGCCATCTCCCATCTTTCTTGAAGAGAAGACTATCA	901
OY	841	GAGAGCATGGAAGAGCAGCTGGGGGAGAGAGGTTTCAGGCCCTCATCCCGCGGACCTTTGCC	900
Db	902	GAGAGCATGGAAGAGCAGCTGGGGGAGAGAGGTTTCAGGCCCTCATCCCGCGGACCTTTGCC	961
OY	901	ATCTCTCAATGCGGAGAGACCGGACTTTCTACCGGCACTGATTACTACACATCCCAATTCCGC	960
Db	962	ATCTCTCAATGCGGAGAGACCGGACTTTCTACCGGCACTGATTACTACACATCCCAATTCCGC	1021
OY	961	CGCCACCTAGACGGTCCCGTCCCGGAGACGGAACCTATCTCGGCGGCATCTCAAGAACACAG	1080
Db	1022	CGCCACCTAGACGGTCCCGTCCCGGAGACGGAACCTATCTCGGCGGCATCTCAAGAACACAG	1081
OY	1021	GAGAAATTAAGACGGCAGCGCCCGTGGGGAGAGAGCGGCTCGGCTGGCTGCGCTCTGCG	1080
Db	1082	GAGAAATTAAGACGGCAGCGCCCGTGGGGAGAGAGCGGCTCGGCTGGCTGCGCTCTGCG	1141
OY	1081	CCGGACATGTTTCCGGAAGCATCTCGCCCGGGTGTACGGCCTGTACCGGCAAGCCCATCTAC	1140
Db	1142	CCGGACATGTTTCCGGAAGCATCTCGCCCGGGTGTACGGCCTGTACCGGCAAGCCCATCTAC	1201
OY	1141	ATCAACCGAAGACGAGTCCCGTGGCTCTTGAAGAGGAAATGATGCTGTGAGAGAGCCGTC	1200

|||||  
Db ATCACCGAAGCAGATGCCGTG-CCTGAGAGAGAAACATGACGTGAGAGAGCCGTC 1260  
QY 1201 AAGACCCCTCCGATCCGCTACTTGACTCGACCTTGACCTGATTTCCAAAGCCATT 1260  
Db 1261 AAGACCCCTCCGATCCGCTACTTGACTCGACCTTGACCTGATTTCCAAAGCCATT 1320  
QY 1261 ACCGAGCAGCGCTGCTGCTCAAGGGGATCTTTCGTCGTCGATTAACCTTGGA 1320  
Db 1221 ACCGAGCAGCGCTGCTGCTCAAGGGGATCTTTCGTCGTCGATTAACCTTGGA 1380  
QY 1321 TGTGTCAGTGGCTACGACCCAGATTGGCGCTCACGTTACAGACTACCACTTCAG 1380  
Db 1381 TGTGTCAGTGGCTACGACCCAGATTGGCGCTCACGTTACAGACTACCACTTCAG 1440  
QY 1381 CGCAGCGCCAGAGTCTGCTGCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440  
Db 1441 CGCAGCGCCAGAGTCTGCTGCTCAAGACATGTTGGCGCCCGCAGAGGTT 1500  
QY 1441 AAGTGGCGCATTA 1455  
Db 1501 AAGTGGCGCATTA 1515

RESULT 3  
AAFI4935  
ID AAFI4935 standard; cDNA; 735 BP.  
XX  
AC AAFI4935;  
XX  
DT 15-SEP-2003 (revised)  
DT 13-MAR-2001 (first entry)  
XX  
DE Trichoderma reesei EST SEQ ID NO:7458.  
XX  
XX Multiple gene expression; filamentous fungal cell; EST;  
KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KM culture condition; environmental stress; spore morphogenesis;  
KM metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
OS Hypocrea jecorina.  
XX  
PN W020056762-A2.  
XX  
PD 28-SEP-2000.  
XX  
PF 22-MAR-2000; 2000WO-US007781.  
XX  
PR 22-MAR-1999; 99US-00273623.  
XX  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
XX  
PI Berka RM, Rey MW, Shuster JF, Kaupinen S, Clausen IG, Olsen PB;  
XX  
DR WPI; 2000-594572/56.  
XX  
PT Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags.  
XX  
PS Claim 89; Page 3012; 3161pp; English.  
XX  
CC The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring the  
CC global expression of genes from FF cells allows the production potential

CC of the microorganisms to be improved. New genes may be discovered,  
CC possible functions of unknown open reading frames can be identified and  
CC gene copy number variation and stability can be monitored. The expression  
CC of genes can be used to study how FF cells adapt to changes in culture  
CC conditions, environmental stress, spore morphogenesis, recombination,  
CC metabolic or catabolic pathway engineering. Using ESTs provides several  
CC advantages over genomic or random cDNA clones including elimination of  
CC redundancy as one spot on an array equals one gene or open reading frame,  
CC and organisation of the microarrays based on function of the gene  
CC products to facilitate analysis of the results. AAF07478 to AAF11247  
CC represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents  
CC ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from  
CC Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
CC Trichoderma reesei, which are all specifically claimed in the present  
CC invention. (Updated on 15-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 735 BP; 168 A; 210 C; 205 G; 148 T; 0 U; 4 Other;

Query Match 41.1%; Score 598; DB 3; Length 735;  
Best Local Similarity 97.0%; Pred. No. 6; Ie-133;  
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

QY 258 CATTCCTCCGCGGCGAGGCTGATCCCGTCAACGAGAGGAAATTGATTTACAGCAA 317  
Db 1 CATTCCTCCGCGGCGAGGCTGATCCCGTCAACGAGAGGAAATTGATTTACAGCAA 60  
QY 318 ACTGATTGACGCGCTGTTGAGGCGGGGTATCAGCGCTTGAGTCTTGTACCTGGGGA 377  
Db 61 ACTGATTGACGCGCTGTTGAGGCGGGGTATCAGCGCTTGAGTCTTGTACCTGGGGA 120  
QY 378 TCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGGCTCAAGTGGAGAGGTTCAGCT 437  
Db 121 TCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGGCTCAAGTGGAGAGGTTCAGCT 180  
QY 438 GAACTTTGACCGGTATGCGAGGTTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTG 497  
Db 181 GAACTTTGACCGGTATGCGAGGTTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTG 240  
QY 498 GATCACCATCAAGAMCCCTGATTCAGGCTCATGTATGATATGCCACCGGAGCAACGC 557  
Db 241 GATCACCATCAAGAMCCCTGATTCAGGCTCATGTATGATATGCCACCGGAGCAACGC 300  
QY 558 CCCGGGAGAGGAGCATTAACAAGACCTTCACCGAGGGCAACCTGCCACTGAGCCGTG 617  
Db 301 CCCGGGAGAGGAGCATTAACAAGACCTTCACCGAGGGCAACCTGCCACTGAGCCGTG 360  
QY 618 GCTGCTGGAAGAGCCCAATCATGAGCCATGCGCGCGCTGGCCGTCTACAGCAGGA 677  
Db 361 GCTGCTGGAAGAGCCCAATCATGAGCCATGCGCGCGCTGGCCGTCTACAGCAGGA 420  
QY 678 CTTTCGCGCCCTGCAAAAGGGCCAGATCGGCATCTGCTCAACGAGGACTACTATGAGCC 737  
Db 421 CTTTCGCGCCCTGCAAAAGGGCCAGATCGGCATCTGCTCAACGAGGACTACTATGAGCC 480  
QY 738 CTGGGACAGCAATGAGCTTCGGGACAGAGAGGCTGCTGAGCGACGAGATGTAATTCAT 797  
Db 481 CTGGGACAGCAATGAGCTTCGGGACAGAGAGGCTGCTGAGCGACGAGATGTAATTCAT 540  
QY 798 TGGCTGTTTGGCAATCCCATCTTTTGAAGAAGACTA-TCCAGAGAGCATGAAGAC 856  
Db 541 TGGCTGTTTGGCAATCCCATCTTTTGAAGAAGACTA-TCCAGAGAGCATGAAGAC 600  
QY 857 AGCT--GGGAGAGGCTTCAGCCCTGAC-TCCGGGAGCTTT--GCCATCTCATGC 911  
Db 601 AACTTGGGGGAGAGGCTTTCAACCTTACTTCCGGGAGCTTTTGGCATCTTMAATGC 660  
QY 912 CGGAGAGAC 921  
Db 661 CGGAGAGAC 670

RESULT 4  
AAFI4882

AAFI4882 standard; cDNA; 1933 BP.  
AAFI4882;  
15-SEP-2003 (revised)  
13-MAR-2001 (first entry)  
Trichoderma reesei EST SEQ ID NO:7405.  
Multiple gene expression; filamentous fungal cell; EST;  
expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
culture condition; environmental stress; spore morphogenesis;  
metabolic pathway engineering; catabolic pathway engineering; ss.  
Hypocrea jecorina.  
WO200056762-A2.  
28-SEP-2000.  
22-MAR-2000; 2000MO-US007781.  
22-MAR-1999; 99US-00273623.  
(NOVO) NOVO NORDISK BIOTECH INC.  
(NOVO) NOVO NORDISK AS.  
Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;  
WPI; 2000-594572/56.  
Monitoring differential expression of genes in filamentous fungal cells  
uses fluorescence-labeled nucleic acids isolated from the cells and a  
substrate of expressed sequence tags.  
Claim 89; Page 2989; 3161bp; English.  
The present invention describes a method for monitoring differential  
expression of genes in a first filamentous fungal (FF) cell relative to  
expression of the same genes in one or more second filamentous fungal  
cells. The method uses fluorescence-labeled nucleic acids isolated from  
the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
are used in the method for monitoring differential expression of genes  
in a first filamentous fungal (FF) cell relative to expression of the  
same genes in one or more second filamentous fungal cells. Monitoring the  
global expression of genes from FF cells allows the production potential  
of the microorganisms to be improved. New genes may be discovered.  
possible functions of unknown open reading frames can be identified and  
gene copy number variation and stability can be monitored. The expression  
of genes can be used to study how FF cells adapt to changes in culture  
conditions, environmental stress, spore morphogenesis, recombination,  
metabolic or catabolic pathway engineering. Using ESTs provides several  
advantages over genomic or random cDNA clones including elimination of  
redundancy as one spot on an array equals one gene or open reading frame,  
and organization of the microarrays based on function of the gene  
products to facilitate analysis of the results. AAF07478 to AAF11247  
represents ESTs from Fusarium venenatum; AAF11248 to AAF11853 represents  
ESTs from Aspergillus niger; AAF11854 to AAF14878 represents ESTs from  
Aspergillus oryzae; and AAF14879 to AAF15337 represents ESTs from  
Trichoderma reesei, which are all specifically claimed in the present  
invention. (Updated on 15-SEP-2003 to standardise OS field)  
Sequence 1933 BP; 467 A; 536 C; 568 G; 343 T; 0 U; 19 Other;  
Query Match 18.3%; Score 266.6; DB 3; Length 1933;  
Best Local Similarity 56.2%; Pred. No. 1.1e-53;  
Matches 770; Conservative 1; Mismatches 540; Indels 59; Gaps 13;  
20 TGCCCAACGACTTTGAATGGGCGCTTCCGACGCGCGCTTACGATGGAAGGCGCGCTCA 79  
250 TGCCCAAGGACTTTCAGTGGGGGTTCCGACGCGCTTCCGATGAGAGGCGCGCTCG 309

80 AAGAGGTGGCCGCGGCGCGCTTCAT--CTGGACAGCTACTGC-----CACTGGAGCCA 132  
ACAGAGACGGCG 310  
TGCGGACCAACG 133  
TGCCCGACGCGCTTCG 370  
GGACTTTGATCTCTTACCAAGTAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 192  
GGAATTTGCGCTCTCAAGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 430  
GCGGATCATTTCCCTCG 252  
GCGCATCATTCG 490  
GAGCAAACTGATTTAGCGCGCTTTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312  
CGTCAAGTTCTGTCGACGACCTGCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 550  
CTGGGATCTGCGCTCAGGCGCGCTTACGATGCTATGAGCGCGCGCGCGCGCGCGCGCG 372  
CTGGGACCTGCG 610  
CCAGCTGACTTTGAGCG 432  
CCGCGCTGACTTTGAAACCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 670  
GAACTGATCACCATCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492  
CAACTGATCACCCTTCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729  
CAAGCG 552  
CTTGCG 789  
GCGCGCTGCTGCTGAAAGCG 612  
GCGGTGACCGTGGCG 819  
GAGGACTT---TGCG 672  
CGACGACTTCAACCG 879  
CTATGAGCG 729  
CACTTACCG 939  
ATTTCACATTTGCTGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 789  
GTTCTTACCG 999  
GAAAGACGACTTGGCG 849  
GCAAGACGACTTGGCG 1058  
TGCCGAGAGACCGGACTTCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 909  
CAAGGCGCTTCAAGGACTTCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1116  
AGAGCGTCCCGTCCCGAGCG 969  
AGCTCG--CCGCTTGGCG 1176  
GAGCGGACG 1029  
GCAAGGCACTGATCG 1234  
ATGTTTCCGAGGATCTCG 1087  
GNTTTTCCGAGGATTTCTGCG 1293  
GAGAGCGGATGCG 1147

Db 1353 GACGAGANACCGNACGAGCATTTAAGGGGAGAGGACCTTGGCAAGAGAAAGATTTNT 1412  
 Qy 1200 CACAGACCCCTTCCGCATCCGCTACTTGGCTGCACTTGGACATCGAT--TTCCAGAGCC 1257  
 Db 1413 CGAAGATGATCTTCAAGGCTCAAGTACTATACAAATACATCCGCTGCAATGGGATACCCGCCG 1472  
 Qy 1258 ATTACCCAGAGACGGCGTGGTCAAGGGGATCTTGGTGGGGCGGTGCTGATTAATCTTG 1317  
 Db 1473 TGGAGCTGAGCGGGGTCAAAAGCTCAAAAGGATCTTGGCTGGTGGCTCATGGACACTTT 1532  
 Qy 1318 GAATGTCAGATGCTTACGAGCCGAGATTCGCGCTCAAGTTTCAAGACTA 1367  
 Db 1533 GAGGGGCGGAGCGGCTACGTAAAGGTTTGGGGTTACTTATGGGGATTA 1582  
 RESULT 5  
 ADA70881  
 ID ADA70881 standard; DNA; 1404 BP.  
 XX ADA70881;  
 AC 20-NOV-2003 (first entry)  
 XX 20-NOV-2003 (first entry)  
 DE Rice gene, SEQ ID 4204.  
 XX Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX Oryza sativa.  
 OS WO2003000898-A1.  
 PN 03-JAN-2003.  
 PD 22-JUN-2001; 2001WO-IB001105.  
 XX 22-JUN-2001; 2001WO-IB001105.  
 PR 22-JUN-2001; 2001WO-IB001105.  
 PA (SYN ) SYNGENTA PARTICIPATIONS AG.  
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
 PI Katsagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 DR WPI; 2003-175290/17.  
 XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX Claim 6; SEQ ID NO 4204; 899bp; English.  
 XX The present invention relates to a method (M) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX Sequence 1404 BP; 356 A; 385 C; 396 G; 266 T; 0 U; 1 Other;  
 Qy Query Match 16.9%; Score 246; DB 7; Length 1404;  
 Db Best Local Similarity 54.5%; Pred. No. 8.7e-49;  
 Matches 725; Conservative 0; Mismatches 565; Indels 40; Gaps 10;  
 62 AGATCGAAGCGCGCTCAAGAAGAGTGGCCGCGCCGCTCATCTGGGACAGCACTGCC 121  
 86 AGTACGAGGAGCGCTGTAAAGAGAGACGCGAGAGGGCGACCACTCTGGGACAGCTTGGCGC 145

Qy 122 ACTGGAACCATGGCGACCAACGCGGCCGATGTGGCTTGGATCTACTACACC 181  
 Db 146 ACACCTTTGAAAGATCACCGACTTTCAGCAATCTGTGTGAGATGATCAGTACACC 205  
 Qy 182 GCTACGATGAGGACTTTGATCTCTTGAACCAAGTACGCGCGCAAGGCTTACCGCTTCCCT 241  
 Db 206 GTTTCGAGGAGATATACAGTCTATGCGAGACATGGGGATGTGATGCGTATCGGTCTTCA 265  
 Qy 242 TGTGTGTGCGCGATATATTCCTCGCGCGAGCGCTGATATCCGCTCAACGAGAGAGAA 301  
 Db 266 TAGATGTGTCAAGATATCTACCAATGTGTG-----TTGGTCAAGTCAATCAAGCTGTGTA 319  
 Qy 302 TTGAGTTTATCAGCAAACTGATATGAGCCCTGTGAGCGGGGTATCAAGCTTGGGTGA 361  
 Db 320 TCGACACATCAACCAAGCTGATGATGCACTTTTACCAAAAGAAATTCAGCCATATGTGA 379  
 Qy 362 CTTTGTACACTGGAGATCTGCTCAGGCGCTTACAGATGCTATGAGGCTGGCTCAACG 421  
 Db 380 CACTTACCACTGGAGCTTCCCGAGCCCTTGAAGACAAATCAAGAGGCTGGCTTGAACA 439  
 Qy 422 TGAAGAGTCCAGCTGGAATTTGAGCGGTATGCGAGTTGTGCTTGAACGTTTGGAG 481  
 Db 440 GG---CAGATAGTGAACGATTTTGGCGCGTACGCGAGAGCTTCAAGGAGTTGGGG 496  
 Qy 482 ACCGAGTCCAGAACTGATATCAATCAACGAMCCCTGATTCAGGCCATCTATGATATG 541  
 Db 497 ACAGGTTGAAGCACTGATATCACTCAACGAGCCGACACGGTGGCATTCAGGGCTAGC 556  
 Qy 542 CCAACCGGACCAAGCGCCCGGAGAGAGACATTA-----ACAAGCACTCCACCGAG 595  
 Db 557 AGCAGAGGGCTCCAGGCGCCCGCGCTGCTCCGTGTGCTCAACCTTATGCAAGCCG 616  
 Qy 596 GCAACACTGCTCACTGAGCCGTGCTGCTGAAAGGCCCAATCATGAGCCATGCCCGC 655  
 Db 617 GCAACTCCGACACCGACCTTACGTGTGCGCCACCACTTACCTCGCCACGCGCCG 676  
 Qy 656 CCGTGGCGCTTACACAGAGGACTTTCGCCCTCGGAAAGGGCCAGATTCGCATTCGC 715  
 Db 677 CCGCCAGCATCTTACAGAGCAAAATACAGGCGACGAGAGCGGCTTGGATATGCT 736  
 Qy 716 TCAACGCGCACTACTATGAGCCCTGGGACGCAATGAGCCTGGGACAGAGAGCTGTG 775  
 Db 737 TCGACGTATGTGTGTGAGCCGATGTCCACACCAAGAT---CGACATTCGGGGCGCA 793  
 Qy 776 AGGAGCGATGGAATTTTCAATTTGCTGTTTGGCAATCCCATCTTCTTGAAGAGAT 835  
 Db 794 AGAGAGCGAGAGATTTTCAAGTATGTTTGTGATCCGTTCTT---CGGCGACT 850  
 Qy 836 ATCCAGAGACATGAAGAAGAGAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGAGAT 895  
 Db 851 ACCCGCGACGATGAGGGGAGAGGCTGGGGAGAGGCTGCCGAGTTCCAGCGGATGAG 910  
 Qy 896 TTGCACTCTCAATGCGGAGAGACCGACTTTCAGCGCATGATTAATACATCACTCCAGT 955  
 Db 911 CGCCCTCTCTAAG---GGGGCGCTGATTTTGTGCGGCTTAACACATCACTCACTACT 967  
 Qy 956 TGGCGGCACTAAGACGCTTCCGTCGCCGANA-----CGGACTATTCGGGCGCATCA 1010  
 Db 968 ACACGAGGACAAACAAACAAACATCATCGGACATTTGCTCAACAAACCTTTGGAGACA 1027  
 Qy 1011 TTAGCAGCAGAGAAAT-----AAGACGCGAGCCCGTTGGCGAGAGAGCGGCTCG 1063  
 Db 1028 CCGGACCGCTCAGCCTCCCATTCAGAAATGGGAACCAATTTGAGATATAGGGAATTTGGA 1087  
 Qy 1064 CCGTGTGCTGCTTCCGCGGACATGTTCCGGAAGCATTCGCGCGGTATAGCGCTGT 1123  
 Db 1088 TATGTGTATCATTTGTGCGCGAGGATGAGAGCGCTGATGAATATATGCAAGAAAGT 1147  
 Qy 1124 ---ACGGCAAGCCCATCTACATCAACGGAACGATGCGCGCTTGGAGAGGAGAA 1180  
 Db 1148 ACACAGGCCACCAAGTATCACTGAAACGGGATGATGACAGCAACACCCGTTCA 1207

Qy	1181	TGACGTGGAGGAGGCGGTCAACACGCCCTTCCGGATCCGCTACTTTGATCCGACCTTGG	1240
Db	1208	TTTTCATTAAGAGACGCCCTTCAAGGACACGAAGAGATTAAATTCACATATACCTTCTCA	1267
Qy	1241	ACTCGATTTCCAGGCCATTACCCAGACGCGCTGTCGTCAAGGGGTACTTTGCGTGGG	1300
Db	1268	CCAAATCTGGCTGCTTCCATCAAGGAGGACGGGTGCAAGTACGTGGGTACTTCCGGTGGT	1327
Qy	1301	CGTTGCTCGATTACTTGGAA-TGGTCAGATGGCTACGGAACCCAGATTTGGCGTCAAGTTT	1359
Db	1328	CTTCGCTGGACAACTGGGAATTTGGGCGCGCCGATATCTCTCGAAGATTCGGGCTCTACTTC	1387
Qy	1360	ACAGACTTACA	1369
Db	1388	GTGGACTTACA	1397

RESULT 6  
ADA69524  
ID ADA69524 standard; DNA; 1503 BP

AC ADA69524;

DT 20-NOV-2003 (first entry)

DE Rice gene, SEQ ID 2847.

KM Plant; bacterial infection; fungal infection; viral infection; rice, KM gene; ds.

**OS** *Oryza sativa*.

PN WO2003000898-A1.

PD 03-JAN-2003

PF - 22-JUN-2001; 2001WO-1B001105.

PR 22-JUN-2001; 2001WO-1B001105.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

DR WPI; 2003-175290/17.

PT Identifying at least one gene involved in plant resistance or response to pathogenic infection for conferring resistance or tolerance to a plant bacterial, fungal or viral infection by determining or detecting plant gene expression.

PS Claim 6; SEQ ID NO 2847; 899pp; English.

The present invention relates to a method (M) for identifying genes involved in plant resistance or response to pathogenic infection. M comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to illustrate the invention.

**SQ** Sequence 1503 BP; 380 A; 420 C; 379 G; 324 T; 0 U; 0 Other;

Query Match	14.7%	Score 214.6	DB 7	Length 1503
Best Local Similarity	51.0%	Pred. No. 2.8e-41		
Matches 720; Conservative	1;	Mismatches 660;	Indels 30;	Gaps 8.

20 TGGCCACGACTTTGATGGGGCTTCGCAAGCGCGCCATCCAGATCGAAGGCGCCGTCA 79

Db	95	TCCCAAGAAATTCATCTTGCGCACAGGTTCAGAGCTTATCAGTATGAGGCGCTTACA	154
Oy	80	AAGAAGTGGCCGGGCCCCGTTCATCTGGGACACGTACTGCCACTGGAGCATTCGGCA	139
Db	155	AAGAAGGGGGGCAAGGCTCTTAGCCCTCTGGGCAACTTCACTCAATTCACAGTAAATTT	214
Oy	140	CCAACGGGGCCCAACGGGAGTGGCTTGGCGATCACTACACCGGTCAGATAGGACTTTG	199
Db	215	TAAACAATGATTAACGGGAGATGTGGCAATGACTTTATCACCGGATACAGAGAGATGGA	274
Oy	200	ATCTCTTGACCAAGTACGGCGCAAAAGGCTACCGCTTCTCCTTGTGTGTGGCGATCA	259
Db	275	GCCTCTGGAAGAGATGAAATGATGATGTTTCGGGTTCTCATTTGGCGGACAGGATCC	334
Oy	260	TTCCCTCGCGCGAGCGTGGATCCCGTCAACGAGAGGGAATTGAGTTTATCAGCAAAAC	319
Db	335	TGCCAAATGGATCCTTGATGAGGAAATAAACAAAGAAAGGGGTGGCTTTCTACAAACGT	394
Oy	320	TGATTTAGCGCCCTTGTAGGGGGGTTACAGCCTTGGGTGACTTTGTACACTGGGATC	379
Db	395	TGAATCAATGATGTATAGCAAAAGGAGATGCCATTTGTACTATCTTTCACATGGGACA	454
Oy	380	TGCTCTCAGGCGCTTACAGATGCTATGAGAGGCTGGCTCAACGTGAAAGAGGTCCAGCTGG	439
Db	455	CCCCCTGGCTCTGGAAGAAACAAATACGAGGATTCCTCA---GTGAAGACTATGTAAGAA	511
Oy	440	ACTTTGACGGTATGCGAGGTTGTGCTTTGAAAGTTTGGGACCGAGTCCAGAACTGGA	499
Db	512	AATACGTGGAATTCGCGAGGGTGTGCTTTCGCGAGTTTCGGGACCGGTGTCAAGTATGGA	571
Oy	500	TCACCATCAACGAAACCCGTGATTCAAGGCATCTATGATATAGCAACGGGACGAACGGCC	559
Db	572	CCACAATTTAATGAGCCATTTCATATAGCGCTACGGCTACGGCAAGGGCGCTGTTTGGCG	631
Oy	560	CGGGCAGGAGCAGCA-----TTAAACAAGCACTCACCGAGGGGCAACCTGCCACTGAGC	613
Db	632	CGGAGCAATGCTCTTCAATATGTTTCCAAGTATGCGGCTGTGGTGACTCCAGTGGCGAGC	691
Oy	614	CGTGGCTGCTGAAAGGCGCCAGATCATGAGCCATGCCCCGGCGGTGGCGCTTACAGCA	673
Db	692	CCTACCTGTGGACACACACATCCACTCTCCACGCTGCGGCTGTCCAGCTTTACCGCA	751
Oy	674	GGAAGCTTTCGCCCTCGCAAAAGGGCCAGATCGGCACTCGCTCAACGGGCGACTATATG	733
Db	752	CCAAGTACCAACCAACAGAAAGGGAACAGATCGGCACTGATGTGGTCAACCCACTGGTTGCG	811
Oy	734	AGCCCTGGGAACAGAAATGAGCTCGGGGCAAGAGAGGCTGAGCGAAGCATGGAATTTC	793
Db	812	TGCCGTACGACAACTCCGACGCT---GACCGTGGCGCTGTGTCCAAACGAGCTTAGACTTCA	868
Oy	794	ACATTGGCTGTTTGGCAATCCCATCTTCTTTGAAAGAAAGACTATCCAGAGACATGAA	853
Db	869	TCATATGGGTGGTTTATGGAACCTTA---TGTGTGATGTGGATTAACCGAGGACATGAGAG	925
Oy	854	AGCAGCTGGCGAAGGCTTTCAGCCTCACTCCCGGGAATTTGCCATCTCTCAATGCGC	913
Db	926	GTTGGGCTCGCAATCGGCTGCGCAAGATTTCAACGCTCGCGATGTGTGAAGGGCT	985
Oy	914	GAGAGACGCACTTCTAAGGCAATGAATTAATCAACATCCAGTTGGCGGCGCAACCTAAGAG	973
Db	986	CCTA---GACTTTCATCGAGTTAATTAATTAACACCACTTACTAGCTTAAAGTATACCGC	1042
Oy	974	GTCCCGTTCGCCG-----GACGGAATTAATCTGGCGGCCATCATGAGCACCAGAGAAATA	1022
Db	1043	CGCCTAATCTCAAGAGAGTATCTTAAGCACTTCGACAAACCGGCGCAACACACCGGCTTCC	1102
Oy	1028	AGAAGCGGACCCCGTTGGCGAGAGAGACGGCTTCGGCTGGCTGCGCTCTGCGCGACA	1087
Db	1103	GTAATGGCAAAACCATGTGCTCAACAGGAATTTACACGATCTTCTTCAACATACCTCCAG	1162
Oy	1088	TGTTTCGGAACATCTGGCCCGGGGTGTAACGGCCTGTACGGGAAGCCCATTAATCAACCG	1147
Db	1163	GTCCTCGAGACTCTCTCTTCAACCAAGAGAGATACACACCGG---CAATCTATGT	1219

QY 1148 AGAAGCATGCCCCGCTGGAGAGAGAACATGACGCGAGGAGCCGCTCAACGACC 1207  
 DB 1220 TTACGAGAAACGCGATTCATGAGGTAACACAGACACTGCCAGAGCGCTCAAGGATG 1279  
 QY 1208 CTTCCGCATCCGCTTACTTTGATCGACCTTGACCTGCATTTCCAAAGCCATTACCCAGG 1267  
 DB 1280 GACACAGGATGAGTTCACATCAAGCACCTGCAGATTGCTCAACACGCCAT---CAAGA 1336  
 QY 1268 ACGGCGTGTGTCGAAGGGGTACTTTGGCTGGGCTTGTCTGATTAATCTTGAATGTAG 1327  
 DB 1337 ATGGGTTAAGGTGAAGGCTACTTACGCTGACATTCATGACTCTTTGATGGGGTG 1396  
 QY 1328 ATGGCTACGAGCCAGATTCGCGCTCACGTTACAGACTACCAACCTCAAGGCGACAGC 1387  
 DB 1397 ACGGCTACCTTGAGACGTTCCGCTCATCTAGTGCAGCCGCAAGACGCTCAAGCGCTACC 1456  
 QY 1388 CCAGAAATGTGCTCCCTGGTCTCAAGACAT 1418  
 DB 1457 GCAAGGAGTCCAGCTACTGATCGAAGACTT 1487

## RESULT 7

ABZ53130  
 ID ABZ53130 standard; cDNA; 575 BP.

ABZ53130;

28-MAR-2003 (first entry)

Aspergillus oryzae polynucleotide SEQ ID NO 2243.

Aspergillus oryzae; fermentation; fungus; industrial; EST;

expressed sequence tag; gene; ss.

Aspergillus oryzae.

WO200279476-A1.

10-OCT-2002.

22-MAR-2002; 2002WO-1B000890.

30-MAR-2001; 2001JP-00098371.

(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

(NARE-) NAT RES INST BREWING.

(NORQ) NAT FOOD RES INST MIN AGRIC.

Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H,

Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K,

WPI; 2003-046817/04.

Detection of expression of specific *Aspergillus* genes for monitoring the fermentation and growth conditions of the fungus, using DNA probes.

Claim 1; SEQ ID NO 2243; 48bp + Sequence listing; Japanese.

The invention relates to a polynucleotide having any of 6006 specific sequences (ABZ50888-ABZ56893), which are expressed by a fungus under specific culture conditions including one or more of eutrophic, oligotrophic, solid, early germination, alkaline, high temperature, low temperature or maltose culture or polynucleotides stringently hybridizing to these sequences. The polynucleotides are useful for monitoring the progress of fermentation and the growth conditions of a fungus, especially of *Aspergillus oryzae* which is widely used in industrial fermentation. Also monitoring for fungal contamination. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Query Match 14.1%; Score 205.4; DB 7; Length 575;  
 Best Local Similarity 63.0%; Pred. No. 3,4e-39;  
 Matches 317; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

QY 17 CTCTGCCAAGCACTTGAATGGGGCTTGCACAGCGCGCCCTACAGATGGAAGCGCCG 76  
 DB 71 CTCCTCCACCCGACTTTCTATGGGGATTTCCCAACAGGATTTACGATTTGAAGAGCCG 130  
 QY 77 TCAAGAGGTGGCGCGCGCCGCTCCATCTGGGACACGTACTGCCACTGAGCCATCGC 136  
 DB 131 TAAATGAGACGAGACGAGTTCATTCATCTGGGATACATTTGCAAGATCCCGGTAAA 190  
 QY 137 GCACCAAGCGGCAACGCGGATGGCTTGGATCACTACACACCGCTACGATGAGGACT 196  
 DB 191 TTGCTGAGAGCGCCATGGGATGGCTTGGATCACTACATTCATTCGACACAGAGATA 250  
 QY 197 TTGATCTCTTGAACCAAGTACGCGCAAGGCGCTTACCGCTTCTCTTGTGCTGGCGGA 256  
 DB 251 TTGCTTGTCTCAAGGCTTGGGCTTGGCCAGGCAATACGTTTCTTGTCTTGTCCGCA 310  
 QY 257 TCATTCCTCTGCGCGGCGGCTGATCCGCTCAACGAGAGGAATTGATTTTACAGCA 316  
 DB 311 TTATTCACCTCGGTGGCGCAAGACCCCATCATGAGAAAGGCTTGCAATATTATATA 370  
 QY 317 AACTGATGACGCGCTTGTGAGGCGGGGTATCAAGCCTTGGGATTTGTACCACTGGG 376  
 DB 371 AGTTCTGCACGACCTGCATGCTGTGTATCACTCTCTCTGTTACTTTGTTCACCTGG 430  
 QY 377 ATCTGCTCAGGCGCTTCAAGATCGCTATGAGAGCTGTGCTCAACGTAAGAGATCCAG 436  
 DB 431 ATCTTCCGATGAGCTCGACAAACGGCTACGCGGCTCTCCCAATAAGAGATTGCTG 490  
 QY 437 TGAATTTGAGCGGTATGCGAGGTGTGCTTTGAACCTTTTGGGACCGAGTCCAAACT 496  
 DB 491 CAGACTTGGCCACTACGACACGATGTTTCAAGGCTTTCGGCTCGAAAGTTAAGCAAT 550  
 QY 497 GGATCAACATCAACGAMCCCTGG 519  
 DB 551 GGATCACTTCAACGAGCCATGG 573

RESULT 8  
 ABZ13477  
 ID ABZ13477 standard; DNA; 1734 BP.

AC ABZ13477;  
 DT 21-JAN-2003 (first entry)

Arabidopsis thaliana stress regulated gene SEQ ID NO 1282.  
 Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

Arabidopsis thaliana.  
 WO200216655-A2.  
 28-FEB-2002.  
 24-AUG-2001; 2001WO-US026685.  
 24-AUG-2000; 2000US-0227866P.  
 26-JAN-2001; 2001US-0264647P.  
 22-JUN-2001; 2001US-0300111P.  
 (SCRT) SCRIPPS RES INST.  
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 PI Harper JF, Kreps J, Wang X, Zhu T;  
 WPI; 2002-304127/34.





PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.

PS Claim 6, SEQ ID NO 71, 899pp; English.

XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.

XX Sequence 1734 BP; 556 A; 316 C; 385 G; 477 T; 0 U; 0 Other;

Query Match 12.6%; Score 182.8; DB 7; Length 1734;

Best Local Similarity 49.2%; Pred. No. 1.1e-33;  
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

```

Qy 18 TCTGCCAAGCAGCTTTGATGGGCTTCGCAACGGCGGCTTACAGATCGAAGGCGCGT 77
Db 90 TTTTCTGATGATTTTCATCTTTGGACAGCTGCTCGGCTTTCAGTACGAAGTGACAC 149
Qy 78 CAAAGAAGTGGCGCGCGCCGCTCCATCTGGAGACAGTACTGCACTGGAAGCCATCGC 137
Db 150 AAGTGAAGTGGCAAGTCTCCACATATGGGATCATTGACCTTCAGCTTCCAGAAAG 209
Qy 138 CACCAACGGCGCCCAACGGCGATGGCTTGGATCAGTACCAACGGCTTACGATGAGGCTT 197
Db 210 GACCAAAATGATATGATGAGATGTAAGCAATTTTATCATGTTTACAAGATGACAT 269
Qy 198 TGATCTTTGACCAAGTACGGGCGCAAGGCGCTTCTCTGTTGTTGTTGCGCGAT 257
Db 270 AAAATTTGATGAAGAGCTAAACATGAGCGCTTCCGATTTTCAATCTGTTGTAAGTT 329
Qy 258 CATTCCTCCGCGGCGAGCTGATCCGTCACGAGAGGAAATTGATTTCAGCA 317
Db 330 AATACCGAGTGAAGTAAAGATGAGTAAACAAAGAGGTATCAATTTCAAGAGA 389
Qy 318 ACTGATTTGACCGCTTGTGAGCGGGGTATCAACGCTTGGGATCTTTGTACACTGGGA 377
Db 390 TCTCATAGCGAACTTTGCTTAATGACATACCACTTGGATGAGCGCTTATCATTTGGGA 449
Qy 378 TCTGCTCAGCGCTTACAGATCGCTATGAGAGCTGGCTCAACGTGAAGAGTCAAGCT 437
Db 450 CACCCACATCTTTGGAGAGAGATATGTGTGCTTTTAAAGCCCTAAATGTAAG--- 506
Qy 438 GGAATTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGAGCCAGTCCAAACTG 497
Db 507 AGATTTTTCAGATTGTCAGAAATTTGTTTGAAGAGTGTGAAGATTAAGTAATGTG 566
Qy 498 GATCCACATCAAGAACCTTGATTCAGGCATCTATGATATGCAACGGAGCAACGCG 557
Db 567 GACAAATCAACGAACCTTATATATGACTGTTGCGGTTATGATCAAGATTAAGAGGC 626
Qy 558 CCCGGGCAAGAGC-----AGCATTAACAAGCACTCCAGCGGACACACTGCACTGA 611
Db 627 GGCTGAGATGTCTCAAAATGGGTAAAGAAAGTGTACAGGCTGAGATTCAGATCCGA 686
Qy 612 GCCGTGGCTCGCTGGAAGGCCCAATCATGAGCCATGCGCGCGCGCTGCTGACG 671
Db 687 GCTTACATGTTTTCATCATCACTCTTCTGCGCATGCGCTGCGTGAAGAAATTTG 746
Qy 672 CAGGAGCTTTGCGCCCTCGCAAAAGGCGAGATCGCATCTGCTCAAGCGGCACTACTA 731
Db 747 AAAATGTGAAAAAATCTTGGCATATGCGCAAAATTTGGGATATGATCTATCAACAAGATGTT 806
Qy 732 TGAGCCTGAGACGAAGATGAGCTCGGAGCAAGAGGCTGCTGAGCGACGATGGAATT 791

```

```

Db 807 CAGGCTTATCATTTCCGATTTCAACTGACGATTAAGAAACAGCTGAACGAGCTCTTGCCCT 866
Qy 792 TCACATTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAAGACTATCCAGAGCATGA 851
Db 867 TGAAATTTGATGATGATCTTGTGATCAGTCAATTC--ACGAGATTATCCAGAGATTGTAA 923
Qy 852 GAAGCAGCTGGGCGAGAGGCTTCCAGCCCTCACTCCCGCGAGCTTTGCCATCTCAATGC 911
Db 924 AAGTACGGGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 983
Qy 912 CGAGAGACCGAATCTTACGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 971
Db 984 TTCA---TCAGATTGTTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1040
Qy 972 CGGTCCGTCCTCCCGAGACGAGATCTACTCTGGCGCATTCATGA-----GCACCAGA 1022
Db 1041 TCACATTAACCTTGAAGAACTTCTTCAAACTGACACCACTGAGATGGAATGAATGAC 1100
Qy 1023 GAATTAAGACGCGAGCCCGTTGGCGAGAGAGCGGCTGCGCTGCTGCTGCTGCTGCTG 1082
Db 1101 TAATCAGATGATCATATCATGAGCTGAGGAGAAAGGCTTCTTATTTTACACACCC 1160
Qy 1083 GAGCATTTCCGAGACATCTGCGCGGTGTAAGGCTGTAC---GGAAGCCACTTA 1139
Db 1161 GGAAGCTTACGAAAGTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1220
Qy 1140 CATCACCGAAGACGATGCGCGCGCTGCGAGAGAAATGACATGACCTGCGAGAGGCGCT 1199
Db 1221 CATCAAGAAATATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1280
Qy 1200 CAACGACCCCTTCCGATCGGTAATTTGATCTGCACTTGGACTGATTTCCAAAGCCAT 1259
Db 1281 GAAGACACATTTAGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1340
Qy 1260 TACCAGAGAGCGGTGTGCTCAAGGGGTACTTGTGCTGCGCGCTGCTGATTAATTTGA 1319
Db 1341 AGTGAAGATGGGTGATGAGATTAAGAGATTTACGATGTCATTTGATGAGCAATTTTGA 1400
Qy 1320 ATGTCAGATGCTTACGAGCCAGATTGCG 1349
Db 1401 ATGGAGCATGATTAACATGCAAGATTGG 1430

RESULT 10
ADA70381
ID ADA70381 standard; DNA; 1533 BP.
AC ADA70381;
DT 20-NOV-2003 (first entry)
XX
XX Rice gene, SEQ ID 3704.
DE
XX Plant; bacterial infection; fungal infection; viral infection; rice;
KM gene; ds.
XX
XX Oryza sativa.
OS
XX
XX WO200300898-A1.
PN
XX
XX 03-JAN-2003.
PD
XX
XX 22-JUN-2001; 2001WO-IB001105.
PF
XX
XX 22-JUN-2001; 2001WO-IB001105.
PR
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
PA
XX
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
PI
XX
XX MPI; 2003-175290/17.
DR

```



XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.

PS Claim 6; SEQ ID NO 3704; 899bp; English.

XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC the expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.

XX Sequence 1533 BP; 414 A; 365 C; 402 G; 352 T; 0 U; 0 Other;

Query Match 12.3%; Score 178.6; DB 7; Length 1533;

Best Local Similarity 49.3%; Pred. No. 1.1e-32;

Matches 705; Conservative 0; Mismatches 694; Indels 30; Gaps 8;

QY 12 GCTAGCTGCTCCCAAGCACTTTGAATGGGCTTCGCAACGGCCCTACAGATCGAAG 71  
 DB 126 GCGAGCTTCCCGAGAGGGGTTCTGTTCCGACGCGCGCTCGGCGTACAGATGAGG 185  
 QY 72 CGCGCTCAAGAGAGTGGCGCGCGCGCTTCATCTGGAGACGTACTCCAGCTGAGCC 131  
 DB 186 CATGGGAGAGAGGGTGGCGGGGCTTACATCTGGAGCGTTTATAGAGAACAGG 245  
 QY 132 ATGGCGCACCAAGCGGCGCAAGCGATGGCTTGATCACTACACCGCTACATGA 131  
 DB 246 GAGCATCTCTAATATATGCAACAGCCAGTGAAGTGAATATATATATATATATAT 305  
 QY 192 GCACTTGAATCTCTTACCAAGTACGCGCAAGCGCTACCGCTTCTCTTCTGCTG 251  
 DB 306 AAGATGATCAAT 365  
 QY 252 GCGGATCTTCCCTCGCGCGCGAGGCTGATCCGCTCAACGAGAGAGGAATGATTTA 311  
 DB 366 AAGAAATTTTCCCAATATGAGATGGAGTGGATG-----GTGAACAGAGAGGAATGATTTA 419  
 QY 312 CAGCAAACTGATGAGCGCTCTTGAAGCGGGATACCGCTTGGGTGACTTTGATCA 371  
 DB 420 CAACAGGTTAAT 479  
 QY 372 CTGGGATCTGCTCAGCGCGCTTACAGATCGCTATGAGGCTGCTCAACGTGAAGAGT 431  
 DB 480 CTATGACTTACCTTATGAGCACTCCATGAGCACTTATAGGTGCTAAGC---CCAAACAT 536  
 QY 432 CCACTGAGCTTTGAGCGGCTTACGAGGTTGCTTTGAAAGTTTGGGAGCCGAGTCA 491  
 DB 537 CGTAGAGGGGCTTTCAGAGACTACGAGATTTCTCTTCCAGACGTTGAGAGAGGGTGA 596  
 QY 492 GAATCTGATCACTATCAACGAMCCCTGGATTCAGGCTATATAGATATCCACCGGAG 551  
 DB 597 GAGCTGTTTACCTTCAATAGAGCGAGATGCGTTGCTCTTGAATATATATATATAT 656  
 QY 552 CAACGCGCGCGAGAGAGCACTTAAACAGCACTCCACGAGAGGCAACCTGCACTGA 611  
 DB 657 CCAATGCCCGGAGAGGCTCTGCTGCTGATG-----CAGAGAGCACTCCACGAGAGA 710  
 QY 612 GCGGTGCTTCTGAGAAAGGCCAGATCATGAGCCATGCCGCGCTGAGCTTACAG 671  
 DB 711 ACCATATCTTGTGCGCAACCATCTCATCTTCTCATGCTGCGCGCTCAAGGATACCG 770  
 QY 672 CAGGACTTTGGCGCTCGAAAGGCGGAGATCGGATCTCGCTCAACGCGCACTACTA 731  
 DB 771 CGAAAGTATCAAGCTTATCAAAAGGAGGAGATTTGAAATTTCTTGGATTTTCTGTGTA 830

QY 732 TGAGCCCTGGGACAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCGACGATGAAATT 791  
 DB 831 TGAACCATTCAGTGAAGCAATGCT---GATAGGGCTGACAGACAGAGGCAAGATTT 887  
 QY 792 TCATATGGCTGTGTTGCCAATCCATCTTCTTGAAGAGACTATCCAGAGAGCTGA 851  
 DB 888 TCACCTCGAGATGTTCTTGAAGCCCATATATC---CATGTGGTACCCATCTCATGTCT 944  
 QY 852 GAAGCAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGAGACTTTGCCATCTCAATGC 911  
 DB 945 CGAATCTGTCAGAGACAGATGCTCACTTCACTATGAGAGTCCAGATGATGTAAGA 1004  
 QY 912 CGAGAGACCGACTTCTACGCAATGATTAATCAACATCCCACTGCGCGCCACTTGA 971  
 DB 1005 CTCATATGATATATGTTGGATCAACACCTTCTTCTATCATAGAAAGCCCTGGGCG 1064  
 QY 972 CGGTCCGCTCCCGAGAGCGATATCTGCGGCTATCATAGACACAGAGAAATAGG- 1030  
 DB 1065 ATGGAACCTGACACCCAGCAATATCATGATGATGCTGCAATGTTGGATTGCTATGAAAG 1124  
 QY 1031 --ACGGCAGCGCGCTTGGGAGAGAGCGCGCTCGCTGAGCTGCTGCGCCGACAT 1088  
 DB 1125 AAGCGCGTCCCATTTGAGCTTACCAAACTCTCTTATGATTTGCTGATAGGG 1184  
 QY 1089 GTTCCGGAAGCATCTGCGCGCGGTGATACGCGCTGTAACGCAAGCCCATCTACACGA 1148  
 DB 1185 AATCAACAAAGGCTGTGACCTATATATATATATATATATATATATATATATATAT 1244  
 QY 1149 GAAAGATGCGGCTGCTGAGAGAGAAATATGATGCTGAGAGAGCGCTCAACGACCC 1208  
 DB 1245 TGAATAATGATATGAGAC---AACCTGCAACGCTATATATATATATATATATATAT 1301  
 QY 1209 CTTCGCGATCCGCTTATCTTGAATCGCATCTGATGCTGATTTCCAGAGCCATTAACCCAGA 1268  
 DB 1302 AGTAAGATCAATATCT---ACAGAAATCATATCATATGAGCTCAAGAGCGATAGACGA 1358  
 QY 1269 CGCGCTGCTGCTCAAGAGGATCTTTGCGTGGCGTGTCTGATTAATGATGTCAGA 1328  
 DB 1359 TGGTGCCTCAAGATGATATGATATCTTGTGCTTGTGATGATGATGATGAGGCT 1418  
 QY 1329 TGGCTTACGAGCCAGATTTGGCGCTCACGTTACAGACTACACCACTCAAGGCGACGCC 1388  
 DB 1419 CGGATCACTTCCGCTTTGGCATCTCTTCAAGTGAAGCTCAAGACCTTAAGAGGATACC 1478  
 QY 1389 CAAGAATGTCGCGCTGCTCTCAAGGACATGTTTGGGCGCGCGCAGAGG 1437  
 DB 1479 CAAGACTATGCTTCTGTTCAAGAACATGCTCTCAATAGAGAGG 1527

RESULT 11  
 AAS21370  
 ID AAS21370 standard; cDNA, 2016 BP.  
 XX AAS21370;  
 AC  
 XX  
 DT 24-OCT-2001 (first entry)  
 XX  
 DE Human cDNA sequence encoding for PRO9820 polypeptide.  
 XX  
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast;  
 KW prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage;  
 KW ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte;  
 KW A-peptide; factor VIIA; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200140466-A2.  
 XX  
 PD 07-JUN-2001.  
 XX  
 PF 01-DEC-2000; 2000MO-US032678.  
 XX  
 PR 01-DEC-1999; 99MO-US028301.

PR 01-DEC-1999; 99WO-US028634.  
 PR 02-DEC-1999; 99WO-US028551.  
 PR 02-DEC-1999; 99WO-US028564.  
 PR 02-DEC-1999; 99WO-US028565.  
 PR 09-DEC-1999; 99US-0170262P.  
 PR 16-DEC-1999; 99WO-US030095.  
 PR 20-DEC-1999; 99WO-US030011.  
 PR 30-DEC-1999; 99WO-US030999.  
 PR 30-DEC-1999; 99WO-US031243.  
 PR 30-DEC-1999; 99WO-US031274.  
 PR 05-JAN-2000; 2000WO-US000219.  
 PR 06-JAN-2000; 2000WO-US000277.  
 PR 06-JAN-2000; 2000WO-US000376.  
 PR 11-FEB-2000; 2000WO-US003565.  
 PR 18-FEB-2000; 2000WO-US004341.  
 PR 18-FEB-2000; 2000WO-US004342.  
 PR 22-FEB-2000; 2000WO-US004414.  
 PR 24-FEB-2000; 2000WO-US004914.  
 PR 24-FEB-2000; 2000WO-US005004.  
 PR 01-MAR-2000; 2000WO-US005601.  
 PR 02-MAR-2000; 2000WO-US005841.  
 PR 03-MAR-2000; 2000US-0187202P.  
 PR 10-MAR-2000; 2000WO-US006319.  
 PR 15-MAR-2000; 2000WO-US006884.  
 PR 20-MAR-2000; 2000WO-US007377.  
 PR 21-MAR-2000; 2000WO-US007532.  
 PR 30-MAR-2000; 2000WO-US008439.  
 PR 17-MAY-2000; 2000WO-US013705.  
 PR 22-MAY-2000; 2000WO-US014042.  
 PR 30-MAY-2000; 2000WO-US014941.  
 PR 02-JUN-2000; 2000WO-US015264.  
 PR 05-JUN-2000; 2000US-0209832P.  
 PR 28-JUL-2000; 2000WO-US020710.  
 PR 11-AUG-2000; 2000WO-US022031.  
 PR 23-AUG-2000; 2000WO-US023522.  
 PR 24-AUG-2000; 2000WO-US023328.  
 PR 08-NOV-2000; 2000WO-US030952.  
 PR 10-NOV-2000; 2000WO-US030873.  
 XX (GETH ) GENENTECH INC.  
 PA Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;  
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
 PI Smith V, Stewart RA, Tumas D, Watanabe CK, Wood WI, Zhang Z;  
 XX WPI: 2001-408281/43.  
 DR P-PSDB; AAU12298.  
 XX  
 PT Isolated, secretory and transmembrane PRO polypeptide used to detect  
 PT other PRO polypeptides, link bioactive molecules to cells expressing PRO  
 PT polypeptides, and detect the presence of mammalian tumors e.g. lung,  
 PT breast, prostate, cervical.  
 XX  
 XX Claim 3; Fig 253; 813p; English.  
 PS  
 XX AAS1244-AAS21518 encode for novel human secretory and transmembrane PRO  
 CC polypeptides. The PRO polypeptides are useful to detect other PRO  
 CC polypeptides, to link bioactive molecules to cells expressing PRO  
 CC polypeptides, and to detect the presence of mammalian lung, colon,  
 CC polypeptides, and to detect the presence of mammalian lung, colon,  
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO  
 CC polypeptide expression in a cell sample to that in a control sample. Some  
 CC of the 275 sequences are also useful to stimulate the release of tumour  
 CC necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or  
 CC differentiation of chondrocytes, the proliferation or gene expression in  
 CC pericyte cells, the release of proteoglycans from cartilage, the  
 CC proliferation of inner ear utricular supporting cells or of T-  
 CC lymphocytes, the release of a cytokine from peripheral blood monocytes  
 CC (PBMCs), or the proliferation of endothelial cells. Some of the PRO  
 CC polypeptides may modulate glucose or free fatty acid uptake by skeletal  
 CC muscle cells or by adipocytes; or inhibit binding of A-peptide to factor  
 CC VIIA. The PRO polypeptides can be used in assays to identify molecules  
 CC involved in binding interactions. The polynucleotides encoding PRO

CC polypeptides can be used to generate probes, antisense RNA/DNA,  
 CC transgenic or knock out animals and can be used in gene therapy  
 XX  
 SQ Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;  
 Query Match 12.2%; Score 178.2; DB 4; Length 2016;  
 Best Local Similarity 54.5%; Pred. No. 1.5e-32;  
 Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;  
 QY 15 AGCTGCCCCAAGCTTTGAAATGGGGCTTCCCAACGGCCGCTACAGATGAAAGCCG 74  
 DB 237 AACCTTCCTCTTGGCTTCTCTGGGCGCTGGGACATTTCTGCTTACACAGAGGGCC 296  
 QY 75 CGTCAAGAAGGTGCGCGCGCCGCTCATCTGGGACAGTACTGCACTTGGAGCCATC 134  
 DB 297 CTGGAGCAGGAGCGGAAAGGGCTTGACATCTGGGACGCTTACACACAGTGGGAAGG 356  
 QY 135 GCGCACCAACGG--CGCAACGGCGATGTGGCTTGCATCTACCTACCGCTAGATGA 191  
 DB 357 GAAAGTCTTGGGAATGAGACGGCAGATGTAGCTGTGACGGCTACTACAGGTCCAGGA 416  
 QY 192 GGACTTGATCTCTTGACCAAGTACGGCGCAAGGCTACCGCTTCTCTGTGCTGTC 251  
 DB 417 GGAATCATTTCTGCTGAGGGAAGTGAACGTCACACATTCATTCCTCTGTCTGGCC 476  
 QY 252 GCGGATCATTTCCCTCGCGCGGAGGCTGATCCGTCACAGAGAGGAATTAATT 311  
 DB 477 CCGGCTCTGCGCCACAGGACATCCGACCGAGCGAGTGAAACAGAAAGGAATCAATTCTA 536  
 QY 312 CAGCAAACTGATGAGCGCCCTGTTGAGGGGGGTATCAGCCCTTGGTGAATTTATCA 371  
 DB 537 CAGTGATCTTATGATGATGCTTCTTGAGCAGCAACATCTCCATGTGATCTTGACCA 596  
 QY 372 CTGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGCTGAGCTCACTGGAAGAT 431  
 DB 597 CTGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGCTGAGCTCACTGGAAGAT 431  
 QY 432 CAGCTGGAATTTGAGCGGTATGCGAGTTTGTCTTGAACGTTTGGGACCGAGTCCA 491  
 DB 657 CAA---CTACTTCAGAGACTACCCCAACCTGTCTTGAAGCCCTTGGGACCGTGGAA 713  
 QY 492 GAATGATGATCAACATCAAGAGCCGATTCAGGCGCATATGATGATCCACCGGAG 551  
 DB 714 GCATGATGATCAACATCAAGAGCCGATTCAGGCGCATATGATGATCCACCGGAG 551  
 QY 552 CAACGCCCGCGGAGGAGCAGCATTTAACAAGCACTCCACGAGGCAACTGCCACTGA 611  
 DB 774 CCAATGGCGCGGCGCTGAAGCTCCGC-----GGCACCGG 806  
 QY 612 GCCGTGCTGCTGGAAAGGCCAGATCATGAGCCATCCCGCGCGCTGCTACAG 671  
 DB 807 CTTTACAGGACAGCAGCAGCATCATTTAAGCCACGCAAACTTGGCATTTTATTA 866  
 QY 672 CAGGAGCTTTGCGCCCTCGCAAAAGGCGCAGATGCGCATCTCGCTCAACGCGACTA 731  
 DB 867 CACCACTGGCGGAGCAGCAGCATGATGATGATGATGATGATGATGATGATGATGAT 926  
 QY 732 TGAACCTGGAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 791  
 DB 927 GGAACCTGGAGCAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 986  
 QY 792 TCAATTTGCTGCTTTTCCCATTCCTCATCT 820  
 DB 987 CTGCTGGCTGCTGCTTCCCAACCCCATTT 1015  
 RESULT 12  
 ABL88245  
 ID ABL88245 standard; cDNA; 2016 BP.  
 XX ABL88245;  
 AC ABL88245;  
 XX  
 DT 16-MAY-2002 (first entry)



DB 987 CTGCTGGGCTGTTGCCAACCATT 1015

RESULT 13

ABL95734 standard; cDNA; 2016 BP.

ABL95734;

19-JUL-2002 (first entry)

Human angiogenesis related cDNA PRO9820 SEQ ID NO: 347.

Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer; atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder; cerdiant; cytosolic; antiangiogenic; hypotensive; vulnary;

antiarteriosclerotic; gene; ss.

Homo sapiens.

WO200208284-A2.

31-JAN-2002.

09-JUL-2001; 2001WO-US021735.

20-JUL-2000; 2000US-0219556P.

25-JUL-2000; 2000US-0220624P.

25-JUL-2000; 2000US-0220644P.

28-JUL-2000; 2000WO-US020710.

02-AUG-2000; 2000US-00643657.

17-AUG-2000; 2000WO-US023522.

23-AUG-2000; 2000WO-US023328.

24-AUG-2000; 2000US-0230978P.

07-SEP-2000; 2000US-00664610.

18-SEP-2000; 2000US-00665350.

24-OCT-2000; 2000US-0242922P.

08-NOV-2000; 2000US-00709228.

08-NOV-2000; 2000WO-US030952.

10-NOV-2000; 2000WO-US030873.

01-DEC-2000; 2000WO-US032678.

20-DEC-2000; 2000US-00747259.

20-DEC-2000; 2000WO-US034956.

22-JAN-2001; 2001US-00767609.

28-FEB-2001; 2001US-00796498.

28-FEB-2001; 2001WO-US006520.

01-MAR-2001; 2001WO-US006666.

09-MAR-2001; 2001US-00802706.

14-MAR-2001; 2001US-00808689.

22-MAR-2001; 2001US-00816744.

05-APR-2001; 2001US-00828366.

10-MAY-2001; 2001US-00854280.

25-MAY-2001; 2001US-00864280.

25-MAY-2001; 2001US-00866028.

25-MAY-2001; 2001WO-US017092.

30-MAY-2001; 2001US-00870574.

30-MAY-2001; 2001WO-US017443.

01-JUN-2001; 2001WO-US017800.

20-JUN-2001; 2001WO-US019692.

XX (GETH) GENENTECH INC.

PA (BAKE) BAKER K P.

PA (FERR) FERRARA N.

PA (GERB) GERBER H.

PA (GERR) GERRITSEN M E.

PA (GODD) GODDARD A.

PA (GODO) GODOWSKI P J.

PA (GURN) GURNEY A L.

PA (HILL) HILLAN K J.

PA (MARS) MARSTERS S A.

PA (PANJ) PAN J.

PA (PAON) PAONI N F.

PA (STEP) STEPHAN J F.

PA (WATA) WATANABE C K.

PA (WILL) WILLIAMS P M.

PA (WOOD) WOOD W I.

PI Baker KP, Ferrara N, Gerber H, Gertlesen ME, Goddard A, Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF, Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

PI Godowski PJ, Watanabe CK, Williams PM, Wood WI, Ye W;

PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;

XX MPI; 2002-171999/22.

DR P-PSDB; ABB95596.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides, useful in diagnosis and treatment of cardiovascular (e.g. myocardial infarction), endothelial or angiogenic disorders in a mammal.

XX Claim 1; Fig 347; 567BP; English.

XX The present invention provides the protein and coding sequences of human PRO proteins. These are useful for treating or diagnosing a cardiovascular, endothelial or angiogenic disorder, including cardiac hypertrophy, trauma, cancer, age-related macular degeneration, atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumor angiogenesis (such as breast carcinoma and liver carcinoma) and wound healing. The present sequence is a coding sequence of the invention

XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

SQ

Query Match 12.2%; Score 178.2; DB 6; Length 2016;

Best Local Similarity 54.5%; Pred. No. 1.5e-32;

Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

OY 15 AGCTGCGCCCAAGCATTTGAATGGGGCTTGGCAAGCGCGCTTACAGATCGAAGCGC 74

DB 237 AACCTTCCTCTTGGCTTCTCTGGGCGCTGGGCAATTCCTTACAGAGGCGCG 296

OY 75 CGTCAAGAGAGTGGCGCGCGCGCTTCATCTGGGACAGTACTGCCAGCTGAGCCATC 134

DB 297 CTGGGACCAAGAGCGGGAAGGCGCTAGCATCTGGGAGCTTTCACACAGTGGAGAGG 356

OY 135 GCGGACCAAGG--CGCCACCGCATGNGGCTTGGAGTACATCACCGCATGATGA 191

DB 357 GAAAGTCTTGGAAATGAGCGGAGATGAGCTTGAGCGCTTACTCAAGGTCCAGGA 416

OY 192 GGAATTGATCTCTTGAACAAGTACGCGCAAGGCGCTTACCGCTTCTGTCGTGTC 251

DB 417 GACATCATTTCTGTGAGGAACTGCACGTCACACTACGATTCCTCGCTTGGCC 476

OY 252 GCGGATCATTCCTCGGCGCGAGGCTGATCCCGTCAAGAGAGGGAATGATTTTA 311

DB 477 CCGGCTCTGGCCCAAGGCACTCCGAGCGGAGCGAGTGAAGAAGGAATGAAATCTA 536

OY 312 CAGGAATGATTTGAGCGCGCTTGAAGCGGGGATACCGCTTGGGTGACTTTGATCA 371

DB 537 CAGTATCTTATGATGCGCTTCTTGAACAGACATATCTCCATCTGACCTTGGACCA 596

OY 372 CTGGATCTGCTCAGGCGCTTCAAGATCGTATGAGGCTGAGCTCAAGTGAAGGCT 431

DB 597 CTGGATCTGCAAGCGCTGCTCAAGTCAATACGAGTGGGAGATGAGACATGCG 656

OY 432 CCACTGAGCTTTTGAAGCGGATGAGAGTTCGCTTTGAAAGTTTGGGAGCGAGTCA 491

DB 657 CAA---CTACTTCAGAGCTACGCAACCTGTGCTTTGAGGCTTTGGAGCGGTGAA 713

OY 492 GAATGATGATCAATCAAGCAAGCCCTGATTCAGGCAATCATGATGATGACCGGAG 551

DB 714 GCACTGATCACTTCACTGATCTCTCGGCAATGAGCAAGAAAGGCTATGAGAGGGCA 773

OY 552 CAAGCCCCGGGAGAGAGCAAGCATTAACAAGCACTCCACGAGGCAACACTGCACTGA 611

DB 774 CCAATCGCGCGGCGCTGAAGTCTCCG-----GGCACCGG 806

PR	05-OCT-1999	99MO-US0230189
PR	29-NOV-1999	99MO-US028214
PR	30-NOV-1999	99MO-US028313
PR	01-DEC-1999	99MO-US028409
PR	01-DEC-1999	99MO-US028301
PR	01-DEC-1999	99MO-US0286344
PR	02-DEC-1999	99MO-US028551
PR	02-DEC-1999	99MO-US028564
PR	02-DEC-1999	99MO-US028565
PR	16-DEC-1999	99MO-US030095
PR	20-DEC-1999	99MO-US030911
PR	20-DEC-1999	99MO-US030999
PR	22-DEC-1999	99MO-US030720
PR	30-DEC-1999	99MO-US031243
PR	30-DEC-1999	99MO-US031274
PR	05-JAN-2000	2000MO-US000219
PR	06-JAN-2000	2000MO-US000277
PR	05-JAN-2000	2000MO-US000376
PR	11-FEB-2000	2000MO-US000365
PR	18-FEB-2000	2000MO-US004341
PR	18-FEB-2000	2000MO-US004342
PR	22-FEB-2000	2000MO-US004414
PR	24-FEB-2000	2000MO-US004914
PR	24-FEB-2000	2000MO-US005004
PR	01-MAR-2000	2000MO-US005601
PR	02-MAR-2000	2000MO-US005746
PR	02-MAR-2000	2000MO-US005841
PR	17-MAY-2000	2000MO-US001305
PR	10-MAR-2000	2000MO-US006319
PR	15-MAR-2000	2000MO-US006884
PR	20-MAR-2000	2000MO-US007337
PR	21-MAR-2000	2000MO-US007532
PR	30-MAR-2000	2000MO-US008439
PR	12-MAY-2000	2000MO-US001370
PR	21-MAY-2000	2000MO-US001402
PR	30-MAY-2000	2000MO-US014941
PR	02-JUN-2000	2000MO-US015644
PR	28-JUL-2000	2000MO-US020710
PR	11-AUG-2000	2000MO-US022031
PR	23-AUG-2000	2000MO-US023522
PR	24-AUG-2000	2000MO-US023328
PR	08-NOV-2000	2000MO-US030952
PR	10-NOV-2000	2000MO-US030873
PR	01-DEC-2000	2000MO-US032678
PR	20-DEC-2000	2000US-0074259
PR	20-DEC-2000	2000MO-US034956
PR	28-FEB-2001	2001US-00764998
PR	28-FEB-2001	2001MO-US006520
PR	01-MAR-2001	2001MO-US006565
PR	09-MAR-2001	2001US-00802706
PR	14-MAR-2001	2001US-00808689
PR	22-MAR-2001	2001US-00816744
PR	05-APR-2001	2001US-00828366
PR	10-MAY-2001	2001US-00854208
PR	10-MAY-2001	2001US-00854280
PR	18-MAY-2001	2001US-00860216
PR	25-MAY-2001	2001US-00866328
PR	25-MAY-2001	2001MO-US017092
PR	01-JUN-2001	2001US-00872035
PR	01-JUN-2001	2001MO-US017800
PR	05-JUN-2001	2001US-00874503
PR	14-JUN-2001	2001US-00882636
PR	19-JUN-2001	2001US-00886342
PR	20-JUN-2001	2001MO-US019692
PR	21-JUN-2001	2001US-00887879
PR	22-JUN-2001	2001MO-US920216
PR	29-JUN-2001	2001MO-US021735
PR	09-JUL-2001	2001MO-US021306
PR	18-JUL-2001	2001US-00908827
PR	06-AUG-2001	2001US-00924419
PR	09-AUG-2001	2001US-00937796
PR	16-AUG-2001	2001US-00951832
PR	19-DEC-2001	2001US-00028072

XX (GETH ) GENENTECH INC.  
 XX Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W,  
 XX Gerliten ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S,  
 XX Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WL, Zhang Z,  
 XX WPI; 2003-341980/32.  
 XX P-PSDB; ABO17742.  
 XX  
 XX New secreted and transmembrane PRO nucleic acids, for treating  
 XX inflammation, organ failure, atherosclerosis, cardiac injury,  
 XX infertility, birth defects, premature aging, acquired immunodeficiency  
 XX syndrome (AIDS), or cancer.  
 XX  
 XX Claim 2; Fig 253; 660pp; English.  
 XX  
 XX The invention describes an isolated nucleic acid (I) comprising, or which  
 XX has 80 % sequence identity to, or the full-length coding sequence of, one  
 XX of 275 nucleotide sequences, and which encodes a corresponding  
 XX polypeptide selected from 275 amino acid sequences, where all sequences  
 XX are given in the specification. The polypeptide encoded by (I) is used to  
 XX detect PRO polypeptides, link a bioactive molecule to a cell expressing a  
 XX PRO polypeptide, modulate a biological activity of a cell, stimulate the  
 XX release of tumour necrosis factor (TNF)-alpha from human blood, modulate  
 XX the uptake of glucose or free fatty acid by cells, stimulate or inhibit  
 XX the proliferation or differentiation of cells or gene expression,  
 XX stimulate the release of proteoglycans, inhibit the binding of cytokine  
 XX to factor VIIa, or detect the presence of tumour in a mammal. The nucleic  
 XX acid and polypeptide encoded by it, are useful for treating inflammatory  
 XX diseases, organ failure, atherosclerosis, cardiac injury, infertility,  
 XX birth defects, premature aging, acquired immunodeficiency syndrome  
 XX (AIDS), cancer, or diabetic complications. The nucleic acid is useful as  
 XX hybridisation probes, in chromosome and gene mapping, and in generating  
 XX antisense RNA or DNA. The polypeptides are useful as pharmaceuticals,  
 XX diagnostics, biosensors or bioreactors. Both are useful in tissue typing.  
 XX This sequence encodes a novel human secreted and transmembrane PRO  
 XX polypeptide  
 XX  
 XX Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;  
 XX  
 XX Query Match 12.2%; Score 178.2; DB 7; Length 2016;  
 XX Best Local Similarity 54.5%; Pred. No. 1.5e-32;  
 XX Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;  
 OY 15 AGCTTGCCCAACGACCTTGAATGGGCTTGCACACGCGCCCTTACCGATCGAAGGCGC 74  
 DB 237 AACCTTCCCTCTTGCTTCTCTGAGGCGTGCGACGTTCTGCTACACGACGAGGCGCG 296  
 OY 75 CGTCAAGAAGGTGGCCGCGGCCGCTCCATCTGGGACACGTAATGCCACCTGGAGCCATC 134  
 DB 297 CTGGAGACGAGGCGGGAAGGGCCCTAGCATCTGGGACCTTTCACACACACATGGGAAAGG 356  
 OY 135 GCGCACCAACGG--CGCCAACGGGATGTGGCTTGCATCACTACACCGCTACGATGA 191  
 DB 357 GAAAGTGTCTGGGAATGAGACGCGAGATGACCTGTACCGCTACCTACCAAGGTCACAGA 416  
 OY 192 GAACCTTGATCTCTTGAACCAAGTACGCGCAAAAGCCCTACCGCTTCTGCTGTGCTGTC 251  
 DB 417 GGACATCATTTCTGCTGAGGGAATGACGATCAACACATCCGATTTCTCGTCTTGGCC 476  
 OY 252 GCGGATCATTTCCCTTCGCGCGGACAGGCTGATCCGTCACAGAGAGGGAATTTGATTTA 311  
 DB 477 CCGGCTCTGCGCCACAGGATCCGAGCCGACAGGTGAACAAAGGGAATCGAATCTTA 536  
 OY 312 CAGCAAACTGATTGACCGCCCTGTGAGCGGGGATATCAACGCTTGGGTGACTTTGTACCA 371  
 DB 537 CAGGATCTTATCGATGCTCTTCTTGAAGCAGCAATCTCCATCGGACCTTGCACCA 596  
 OY 372 CTGGGATCTGCTCAGGCGCTTACGATCGCTATGAGAGGCTGCACGCGAAGAGGT 431  
 DB 597 CTGGGATCTGCAACAGCTGCTCCAGGTCAATACGCTGGGTGGAGAAATGTAGCATGGC 656

OY 432 CCAGCTGACCTTTGAGCGGATATGAGAGTTGTGCTTTGAACGTTTGGGACCGAGTCCA 491  
 DB 657 CAA---CTACTTACGAAACTACGCAACCTGTGCTTTGAGCCCTTTGGGACCGGTGAA 713  
 OY 492 GAACGTGATCAACATCAACGAGCCCTGATTCAGGCCATATGATATGCCACCGGACG 551  
 DB 714 GCACGTGATCAACGTTGATGATGATCCTCGGCAATGGCGAATAAAGGATATGAGACGGGCCA 773  
 OY 552 CAAGCCCCGGGCGAGGACGACATTTACAAAGCACTCCACGAGGGCAACCTGCACCTGA 611  
 DB 774 CCATGCGCCGGGCTGAAAGCTTCGCG-----GGCACGG 806  
 OY 612 GCCGTGCTCGCTTGAAGGCGCCAGATCATGAGCCATGCCGCGCTGCGCTTACAG 671  
 DB 807 CTGTACAAAGGACGACACACCAATCTTAAGGCCACGCAAAACCTGGAATCTTATTA 866  
 OY 672 CAGGACCTTTCGCCCCCTCGCAAAAGGCGCATTCGCTCAACGCGACCTACTA 721  
 DB 867 CACCACGTGCGCGAGCAAGCAGAGAGGTCTGTGGGAATTTCACTGAACCTGTGACGTGGG 926  
 OY 732 TGAGCCCTGGAGACGAATAGCTCTGGGACAAAGAGGCTGTGAGGACGATGGAATT 791  
 DB 927 GGAACCTGTGACATTTGTAATCCCAAGACCTTAGAGGCTGCCGAGATACCTACAGTT 986  
 OY 792 TCACATTGGCTGTTTGCATTCCTCATCT 820  
 DB 987 CTGTCTGGGCTGTGTTTCCCAACCCATTT 1015  
 XX  
 XX RESULT 15  
 XX ACA67120  
 XX ID ACA67120 standard; cDNA; 2016 BP.  
 XX AC  
 XX ACA67120;  
 XX DT 23-JUN-2003 (first entry)  
 XX XX  
 XX cDNA encoding human PRO polypeptide #127.  
 XX  
 XX Human: PRO polypeptide; secreted and transmembrane protein;  
 XX anti-PRO antibody; diagnostic assay; gene expression; diabetes;  
 XX bone disorder; cartilage disorder; rheumatoid arthritis; obesity;  
 XX sports injury; osteoarthritis; hyper-insulinaemia; hypo-insulinaemia;  
 XX hearing loss; coagulation disorder; stroke; heart attack; cardiac;  
 XX antidiabetic; anorectic; vulnery; antiarthritic; osteopathic;  
 XX antipneumatic; auditory; cerebroprotective; angiogenic; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX OS  
 XX US200304311-A1.  
 XX PN  
 XX 02-JAN-2003.  
 XX PD  
 XX  
 XX PF 19-DEC-2001; 2001US-00028072.  
 XX XX  
 XX 18-JUN-1997; 97US-0049911P.  
 XX PR 26-AUG-1997; 97US-0056974P.  
 XX PR 17-SEP-1997; 97US-0059113P.  
 XX PR 17-SEP-1997; 97US-0059115P.  
 XX PR 17-SEP-1997; 97US-0059117P.  
 XX PR 17-SEP-1997; 97US-0059122P.  
 XX PR 17-SEP-1997; 97US-0059184P.  
 XX PR 18-SEP-1997; 97US-0059163P.  
 XX PR 19-SEP-1997; 97US-0059352P.  
 XX PR 19-SEP-1997; 97US-0059388P.  
 XX PR 24-SEP-1997; 97US-0059636P.  
 XX PR 17-OCT-1997; 97US-0062250P.  
 XX PR 17-OCT-1997; 97US-0062285P.  
 XX PR 17-OCT-1997; 97US-0062287P.  
 XX PR 24-OCT-1997; 97US-0063755P.  
 XX PR 24-OCT-1997; 97US-0062814P.  
 XX PR 24-OCT-1997; 97US-0062816P.

PR 24-OCT-1997; 97US-0063045P.  
 PR 24-OCT-1997; 97US-0063082P.  
 PR 24-OCT-1997; 97US-0063127P.  
 PR 27-OCT-1997; 97US-0063327P.  
 PR 27-OCT-1997; 97US-0063329P.  
 PR 28-OCT-1997; 97US-0063550P.  
 PR 28-OCT-1997; 97US-0063561P.  
 PR 29-OCT-1997; 97US-0063704P.  
 PR 29-OCT-1997; 97US-0063733P.  
 PR 29-OCT-1997; 97US-0063735P.  
 PR 29-OCT-1997; 97US-0063738P.  
 PR 03-NOV-1997; 97US-0064248P.  
 PR 07-NOV-1997; 97US-0064809P.  
 PR 12-NOV-1997; 97US-0065186P.  
 PR 17-NOV-1997; 97US-0065846P.  
 PR 21-NOV-1997; 97US-0066364P.  
 PR 24-NOV-1997; 97US-0066453P.  
 PR 24-NOV-1997; 97US-0066511P.  
 PR 24-NOV-1997; 97US-0066707P.  
 PR 11-DEC-1997; 97US-0069212P.  
 PR 11-DEC-1997; 97US-0069334P.  
 PR 16-DEC-1997; 97US-0069694P.  
 PR 23-JAN-1998; 98US-0072320P.  
 PR 04-FEB-1998; 98US-0073612P.  
 PR 09-FEB-1998; 98US-0074082P.  
 PR 12-MAR-1998; 98US-0077791P.  
 PR 20-MAR-1998; 98US-0078910P.  
 PR 25-MAR-1998; 98US-0079294P.  
 PR 27-MAR-1998; 98US-0079663P.  
 PR 27-MAR-1998; 98US-0079728P.  
 PR 31-MAR-1998; 98US-0080165P.  
 PR 12-JUN-1998; 98MO-US012456.  
 PR 14-JUL-1998; 98MO-US014552.  
 PR 28-AUG-1998; 98MO-US01788P.  
 PR 10-SEP-1998; 98MO-US018824.  
 PR 14-SEP-1998; 98MO-US019093.  
 PR 14-SEP-1998; 98MO-US019094.  
 PR 14-SEP-1998; 98MO-US019177.  
 PR 16-SEP-1998; 98MO-US019330.  
 PR 17-SEP-1998; 98MO-US019437.  
 PR 07-OCT-1998; 98MO-US021141.  
 PR 29-OCT-1998; 98MO-US022991.  
 PR 29-OCT-1998; 98MO-US022992.  
 PR 20-NOV-1998; 98MO-US024855.  
 PR 01-DEC-1998; 98MO-US025108.  
 PR 05-JAN-1999; 99MO-US000106.  
 PR 08-MAR-1999; 99MO-US005028.  
 PR 10-MAR-1999; 99MO-US005190.  
 PR 20-APR-1999; 99MO-US008615.  
 PR 14-MAY-1999; 99MO-US010733.  
 PR 02-JUN-1999; 99MO-US012252.  
 PR 01-SEP-1999; 99MO-US020111.  
 PR 08-SEP-1999; 99MO-US020594.  
 PR 13-SEP-1999; 99MO-US020944.  
 PR 15-SEP-1999; 99MO-US021090.  
 PR 15-SEP-1999; 99MO-US021547.  
 PR 05-OCT-1999; 99MO-US023089.  
 PR 29-NOV-1999; 99MO-US028214.  
 PR 30-NOV-1999; 99MO-US028313.  
 PR 30-NOV-1999; 99MO-US028409.  
 PR 01-DEC-1999; 99MO-US028301.  
 PR 01-DEC-1999; 99MO-US028634.  
 PR 02-DEC-1999; 99MO-US028551.  
 PR 02-DEC-1999; 99MO-US028564.  
 PR 02-DEC-1999; 99MO-US028565.  
 PR 16-DEC-1999; 99MO-US030095.  
 PR 20-DEC-1999; 99MO-US030911.  
 PR 20-DEC-1999; 99MO-US030939.  
 PR 30-DEC-1999; 99MO-US031243.  
 PR 30-DEC-1999; 99MO-US031274.  
 PR 05-JAN-2000; 2000MO-US000219.

PR 06-JAN-2000; 2000MO-US000277.  
 PR 06-JAN-2000; 2000MO-US000376.  
 PR 11-FEB-2000; 2000MO-US000365.  
 PR 18-FEB-2000; 2000MO-US000341.  
 PR 18-FEB-2000; 2000MO-US000342.  
 PR 22-FEB-2000; 2000MO-US00414.  
 PR 24-FEB-2000; 2000MO-US004914.  
 PR 24-FEB-2000; 2000MO-US005004.  
 PR 01-MAR-2000; 2000MO-US005601.  
 PR 02-MAR-2000; 2000MO-US005746.

(GETH ) GENENTECH INC.

PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W,  
 PI Gertlsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;  
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

DR WPI; 2003-352836/33.  
 DR P-PSDB; ABU80996.

PT New isolated PRO polypeptide useful for treating diabetes, rheumatoid  
 PT arthritis, sports injuries, obesity, hearing loss in mammals, stroke, or  
 PT heart attack.

Claim 2; Fig 253; 643pp; English.

CC The present invention relates to the isolation of novel human PRO  
 CC polypeptides, and the polynucleotide sequences encoding them. The PRO  
 CC polypeptides are secreted and transmembrane proteins. The PRO  
 CC polypeptides and polynucleotides are useful for preparing a medicament  
 CC useful in the treatment of diabetes, bone and/or cartilage disorders  
 CC (e.g. rheumatoid arthritis, sports injuries, osteoarthritis), obesity,  
 CC hyper- or hypo-insulinaemia, hearing loss, and coagulation disorders  
 CC (e.g. stroke, heart attack). Anti-PRO antibodies are useful in diagnostic  
 CC assays for PRO, by detecting its expression in specific cells, tissues or  
 CC serum, and for affinity purification of PRO from recombinant cell culture  
 CC or natural sources. ACA65994-ACA67268 represent cDNA sequences encoding  
 CC the human PRO polypeptides of the invention. Note: The sequence data for  
 CC this patent was obtained in electronic format directly from the USPTO web  
 CC site at [seqdata.uspto.gov/patseqidentry.html](http://seqdata.uspto.gov/patseqidentry.html)

Sequence 2016 BP; 509 A; 499 C; 516 G; 492 T; 0 U; 0 Other;

Query Match 12.2%; Score 178.2; DB 7; Length 2016;  
 Best Local Similarity 54.5%; Pred. No. 1.5e-32;  
 Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

QY 15 AGCTTGCCCAAGCACTTGAAATGGGGCTTGCAGACGGCCGCTTACAGATCGAAGCCG 74  
 DB 237 AACCTTCCCTCTGGCTTCTCTGGGGCGTGGGACATTTGCTTACAGACGAGGCGC 296  
 QY 75 CGTCAAGAAAGTGGCGCGCCGCTTCATCTGGGACAGCTACTGCACTTGAGCCATC 134  
 DB 297 CTGGGACCAAGGAGGAGGAGGCTTAGCATCTGGGAGCTTTACACAGTGGGAGGG 356  
 QY 135 GCGCACAACGG--GCGCAAGCGATGAGTGGCTTGCGATCACTTACACCGCTAGATGA 191  
 DB 357 GAAAGTCTTGGAAATGAGACGCAATGAGCTGTGACGCTTACTACAGATCCAGGA 416  
 QY 192 GGAATTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTTCTGCTGTC 251  
 DB 417 GGAATCATTTCTGCTGAGGGAATGCAACGTCACATACCACTACGATTCCTCTGCTGGCC 476  
 QY 252 GCGGATCATTCCTCTGGCGGCGAGGCTGATCCCGTCACAGAGAGGAATTAAGTTTA 311  
 DB 477 CCGGCTCTCGCCACAGGATCCGAGCCGAGCGAGGAGGAAACAAAGAGATCAATTTTA 536  
 QY 312 CAGCAACTGATGAGCGCCCTGTTGAGGGGGGATACAGCCCTTGGGTGACTTTGACCA 371  
 DB 537 CAGTGATCTTATGATGACGCTTCTGAGCAGCAACATCACTCCATGATGACTTGCACCA 596  
 QY 372 CTGGGATCTGCTCAGAGCGCTTACAGATCGCTATGAGAGGCTGCTCAGCTGAGAGGT 431

```
Db 597 CTGGATCTGCCACAGCTGCTCCAGTCAAAATACGGTGGGTGGCAAAATGTAGCATGGC 656
QY 432 CCAGCTGGACTTTGAGCCGATATGCGAGTTGTGCTTTGAAAGTTTGGGGACCGAGTCCA 491
Db 657 CAA---CTACTTCAAGAGACTAGCCACCTGTGCTTTAGGGCTTTGGGGACCGTGA 713
QY 492 GAACTGATCAACATCAACGAWCCCTGATTCAGGCCATCTATGATATGCCACCGGAG 551
Db 714 GCACGTGATCAAGTTCAGTGATCTCGGGCAATGGCAGAAAAAGGCTATGAGACGGGCCA 773
QY 552 CAACGCCCCGGGCAAGAGCAGATTAACAAGCACTCCACCGAGGGCAACACTGCCACTGA 611
Db 774 CCATGCGCGGGGCTGAAGCTCCGC-----GGCACCGG 806
QY 612 GCCGTGCTCGCTGGAAGAGCCAGATCATGAGCCATGCCCCGCGCGTGTCTACAG 671
Db 807 CTTGTACAAGGCGAGCACAACACATCATTAAGGCCACGCAAAACCTTGCAATTCTTATTA 866
QY 672 CAGGACTTTGCCCCCTGCAAAAGGGCCAGATCGGCATCTGCTCAACGGCGACTACTA 731
Db 867 CACCACTGGCGCAGCAAGCAGCAAGTCTGGTGGAAATTCACTGAACGTGACTGGGG 926
QY 732 TGAAGCCCTGGGACAGCATGAGCCCTCGGGAACAAGAGGCTGTGAGCGACGATGAAAT 791
Db 927 GGAACCTGTGACATTAAGTAACCCCAAGAGCTTAAGAGGCTGCCGAGAGATACTACAGTT 986
QY 792 TCACATTTGGCTGTTTGGCAATCCCATCT 820
Db 987 CTGTCTGGGCTGTGTTGCCAACCCCATTT 1015
```

Search completed: March 28, 2004, 17:03:18  
Job time : 587.342 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 20:52:22 ; Search time 105.557 Seconds

(without alignments)  
7649.443 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455  
Sequence: 1 atgcccgagtcgctagctc.....gggttaagtgcgcgcataa 1455Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapect 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	10.3	1521	US-09-914-841A-14	Sequence 14, Appl
2	139.4	9.6	1467	US-09-489-039A-2386	Sequence 2386, Ap
3	130.8	9.0	1317	US-09-134-078-2	Sequence 2, Appl
4	128.2	8.8	3163	US-09-344-510B-6	Sequence 6, Appl
5	128.2	8.8	3435	US-09-344-510B-7	Sequence 7, Appl
6	127.8	8.8	1650	US-09-344-510B-9	Sequence 9, Appl
7	127.8	8.8	5032	US-09-344-510B-8	Sequence 8, Appl
8	126.2	8.7	1287	US-09-914-841A-2	Sequence 2, Appl
9	124.8	8.6	1542	US-09-122-230-6	Sequence 6, Appl
10	124.8	8.6	1909	US-09-122-230-5	Sequence 5, Appl
11	118.6	8.2	1829	US-08-688-988-9	Sequence 9, Appl
12	111.4	7.7	1545	US-08-400-275-17	Sequence 17, Appl
13	104.8	7.2	1458	US-09-134-000C-2890	Sequence 2890, Ap
14	101.4	7.0	1437	US-09-107-532A-1268	Sequence 1268, Ap
15	95.6	6.6	1410	US-09-489-039A-1053	Sequence 1053, Ap
16	93.8	6.4	1407	US-08-688-988-7	Sequence 7, Appl
17	90.4	6.2	2435	US-09-306-593-1	Sequence 1, Appl
18	88.4	6.1	1931	US-09-431-470-1	Sequence 1, Appl
19	88.4	6.1	1931	US-09-431-470-3	Sequence 3, Appl
20	87.6	6.0	454	US-09-615-192A-166	Sequence 166, App
21	87.6	6.0	454	US-09-169-789-166	Sequence 166, App
22	85.2	5.9	457	US-08-975-316-79	Sequence 79, Appl
23	85.2	5.9	457	US-09-615-192A-79	Sequence 79, Appl
24	85.2	5.9	457	US-09-169-789-79	Sequence 79, Appl
25	85.2	5.9	470	US-09-615-192A-141	Sequence 141, App
26	85.2	5.9	470	US-09-169-789-141	Sequence 141, App
27	84.8	5.8	3460	US-09-344-510B-10	Sequence 10, Appl

C	28	84.2	5.8	1404	4	US-09-489-039A-4718	Sequence 4718, Ap
	29	84.2	5.8	1455	4	US-09-489-039A-4800	Sequence 4800, Ap
	30	79.4	5.5	1443	4	US-09-489-039A-1986	Sequence 1986, Ap
C	31	74.8	5.1	16535	4	US-08-961-527-74	Sequence 74, Appl
	32	71	4.9	1428	4	US-09-489-039A-2983	Sequence 2983, Ap
	33	69.4	4.8	1479	4	US-09-134-000C-289	Sequence 289, Appl
	34	68.4	4.7	286	4	US-09-313-294A-7321	Sequence 7321, Ap
C	35	68	4.7	3615	1	US-08-920-812-17	Sequence 17, Appl
	36	68	4.7	3615	1	US-08-920-827-17	Sequence 17, Appl
C	37	68	4.7	3615	1	US-08-362-577C-17	Sequence 17, Appl
	38	68	4.7	3615	2	US-08-920-828-17	Sequence 17, Appl
C	39	68	4.3	11303	4	US-08-961-527-115	Sequence 115, App
C	40	63	4.0	1371	4	US-09-134-000C-1598	Sequence 1598, Ap
	41	57.6	4.0	1464	4	US-09-489-039A-4740	Sequence 4740, Ap
	42	56.4	3.9	1464	4	US-09-489-039A-2778	Sequence 2778, Ap
	43	54.6	3.8	1404	4	US-09-489-039A-2724	Sequence 2724, Ap
	44	54.2	3.7	1455	4	US-09-615-192A-138	Sequence 138, App
	45	54	3.7	424	4	US-09-615-192A-138	Sequence 138, App

## ALIGNMENTS

```
RESULT 1
US-09-914-841A-14
; Sequence 14, Application US/09914841A
; Patent No. 6645750
; GENERAL INFORMATION:
; APPLICANT: Amano Enzyme Inc.
; TITLE OF INVENTION: Beta-PRIMEVEROSIDASE GENE
; FILE REFERENCE: 066072
; CURRENT APPLICATION NUMBER: US/09/914,841A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: P. Hei. 11-056299
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/JP00/01242
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Camellia var. sinensis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1521)
; OTHER INFORMATION:
; NAME/KEY: mat peptide
; LOCATION: (235)..()
; OTHER INFORMATION:
US-09-914-841A-14

Query Match      10.3%; Score 150; DB 4; Length 1521;
Best Local Similarity 52.2%; Pred. No. 3.3e-29;
Matches 455; Conservative 1; Mismatches 401; Indels 15; Gaps 5;

QY      20  TGGCCACGACCTTTGAATGGGGCTTGGCAAGCGCCGCTTACGATCGAAGGCGCGTCA 79
DB      116  TTCTGATGTTTGTCTTGGAGCTGCTCTTCTGCTTACGAGTTGAAGTGCTGCCA 175
QY      80  AAGAGGTGGCGCGCGCGCTGCATCGGAGACGTAAGTCCGACCTGCGGCA 139
DB      176  AAGAGGTGGAGAAAGGCCCAATTTTGGATTCCTTCACTCAATGAGTTTCCAGTAAA 235
QY      140  ---CCAAAGCGCCCAAGCGGATGGCTTGCATCACTCAACCGCTAGATGAGACT 196
DB      236  TATCGATGTGTACACTCGAGATGTAGCTGATACCTTTATCAATCGTTAACAGGAAGATG 295
QY      197  TTGATCTCTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCTGTGTGCGGA 256
DB      296  TGAAGGTGCTGAAGTTTATGAGACTAGATGTTTCAGAAATGCTCATCTGAGGCGGAG 355
QY      257  TCATTCCTCCGCGGCGGAGCTGATCCGTCACGAGGAGGAATGATTTACAGCA 316
```

Db 356 TATTACCTGGGGAAGCTTAGCGAGGAGTGAACAAGAGATGCTTCTTACACA 415  
Qy 317 AACTGATTGAGCCCTGTTGAGCGGGTATACAGCCCTTGGGTGACTTTGACACTGGG 376  
Db 416 ATGCATCAATGACCTTTATTCGAAGGTATACCACTTTTATTAACAATCTTTCACTGGG 475  
Qy 377 ATTCGCTCAGGCGCTTCAGCATGCGCTATGAGAGGTGGCTCAACGTGGAAGAGGCCAGC 436  
Db 476 ATCTTCCCAAGCCCTAGAAATGAAATATGAGAGCTTTTAAAGC---CCACACTTTGGA 532  
Qy 437 TGACCTTGAACCGGATGCGAGGTGTGCTTTGAAACGTTTGGGAGCCGAGTCCAGAACT 496  
Db 533 ACGATTTCCGGGATTTTGCAGAGCTGTGCTTCAAGAGATTTGGTGAACGATTTAAACTT 592  
Qy 497 GGATCACCATCAAGAACCTTGATTCAGGCACTATGATATGCCACCGGACGAACG 556  
Db 593 GGATCACAATAAAGAACCATGCTTACTCTTATGGGGTTATGATGACAGTCTCTAG 652  
Qy 557 CCCCCGGCAGAGAGCATTAACAAGCACTC---CACGAGGGGCAACATGCACTGAGC 613  
Db 653 CACGGGCGCTTGTTCGCTTTATGSCATTTTGCCTTAAAGGAAATCTGGAGCTAGC 712  
Qy 614 CGTGACCTGCTGAAAGGCCCAATCATAGCACTGCCGCGCGTGGCCGTCTACAGCA 673  
Db 713 CCTATATAGTAAACCAATTTGCTTTCTCAAGCTGCTGCTGTGAACCTATACAAAG 772  
Qy 674 GGGACTTGGCCCTTCGCAAAAGGCCAGATCGCATCTCGCTCAAGCGGCACTACTATG 733  
Db 773 AGAATATACAGGCATATCAAAAGGGCAGATAGGATTAACACTGTGACTTATGATGA 832  
Qy 734 AGCCCTGGGACAGCAATGAGCCTCGGACAAAGAGGCTGTGAGGACGATGGAATTC 793  
Db 833 TTCCCT---ACTCCAATTCGAAAGCCGACAAAGATGACGACAAAGACCCCTGATTTCA 889  
Qy 794 ACATTTGCTGTGTTGCCAATCCATCTTTGAAGAAAGACTATCCAGAGCATGAAGA 853  
Db 890 TGTATGATGATGTTTATGAGCATTAAGCTT---TGTGATATATCAAAAGACATGCGTA 946  
Qy 854 AGCAGCTGGGCGAGAGGCTTCCAGCCCTCACT 885  
Db 947 GACTCGTTGTGTAAGGTACCTTACCAAGTTCACT 978

RESULT 2  
US-09-489-039A-2386  
; Sequence 2386, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489.039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2386  
; LENGTH: 1467  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2386

Query Match 9.6%; Score 139.4; DB 4; Length 1467;  
Best Local Similarity 56.1%; Pred. No. 1.8e-26;  
Matches 307; Conservative 1; Mismatches 227; Indels 12; Gaps 2;

Qy 16 GCTCGCCCAAGCATTTGAATGGGGCTTGCACAGCGCGCTACCAATGGAAGGGGCC 75  
Db 103 GCTTTCCGCAATTTCTTATGGGGTGGCGGTACCGAGGTATCAGTTGAGGAGGA 162  
Qy 76 GTCAAGAAGGTGGCGGCGCGCTCATCTGGGACAGTACTGCACTGAGGCATCG 135

Db 163 CAGAGCCGACCGGTAAAGAACCGTCAATCTGGGATATCTACTCTCATCTGCCGGTACC 222  
Qy 136 CGACCAACAGCGCCCAACCGCGATGTGGCTTGGATATCAACACCGCTACGATGAGAC 195  
Db 223 ACCTTTGAAGGCAACACCGGCGATATTTGCGTGACACATATCACCGTTTCCGAAGAC 282  
Qy 196 TTTGATCTCTGACCAATGACGGGCAAAAGGCTTCTCTTGTCTGTGGTCCGG 255  
Db 283 GTGGCGTTAATGCGCGAGATGGGCTGCAAAAGCTATGATTCATCTCTGGCCCCG 342  
Qy 256 ATCATTCCTCGCGCGAGCTTGATCCGCTCAACGAGAGGAAATGAGTTTACAGC 315  
Db 343 CTGCTGCC-----CGCCGGCGCGGTAAAGTGAATGAACGGGAGTCCAGTTTACAGC 396  
Qy 316 AAATGATTAAGCCCTGTTGAGCGGGGATATCACGCTTGGGTGATTTGTACACTGG 375  
Db 397 GATCTGATGACAGCAACTGTGGCGCACAAATATGAGCCGATGATCACTCTATCACTGG 456  
Qy 376 GATCTGCTCAGGCGCTTACGATCGCTATGAGGCTGCTCAACGTGAAGAGTCCAG 435  
Db 457 GATCTGCGGAGGCGCTTGCAGG-----ATGAGGCGCGCTGGGAAGCGGCACTACCGCC 510  
Qy 436 CTGACTTTGAGCGGTATGCGAGGTGTGCTTTGAACGTTTGGGACCGAGTCCAGAAC 495  
Db 511 GAGGCTTCCCGAGTACGCGCCGCTGTGCTATGCGCGCTTTGGTTCACGGGTGAACCTG 570  
Qy 496 TGTATCACTATCAAGAACCCCTGATTCAGGCCATCTATGAGATATGCCACCGGAGAAC 555  
Db 571 TGGGCTACTTCAACGAACCATTTGTTCAATTGGGACGGCTACATTAAGGCTTCAT 630  
Qy 556 GCCCGCG 562  
Db 631 CCCCCG 637

RESULT 3  
US-09-134-078-2  
; Sequence 2, Application US/09134078  
; Patent No. 6368844  
; GENERAL INFORMATION:  
; APPLICANT: Bylina, Edward J.  
; TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
; STREET: 4365 Executive Drive, Suite 1600  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/134,078  
; FILING DATE: 13-AUG-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/949,026  
; FILING DATE: 10-OCT-1997  
; APPLICATION NUMBER: 60/056,916  
; FILING DATE: 06-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Halie, Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09010/024002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 858/677-1456  
; TELEFAX: 858/677-1465  
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1317 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1314  
US-09-134-078-2

Query Match 9.0%; Score 130.8; DB 4; Length 1317;  
Best Local Similarity 56.4%; Pred. No. 2.8e-24;  
Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

16 GCCTCCCAACAGCATTTGAATGGGCTTGGCAACGGCCCTTACCAAGATCGAAGGCC 75  
16 GATTTTCAGAAAGATTTTATCTTCGAAACGGCTACGACATACAGATTGAAGTGCA 75  
76 GTCAAGAAAGGTGGCCGGCCCGCTCATCTGGGACAGTACGCACTGAGCCATCG 135  
76 GCAGAGAGAGATGGCAGAGGCGCATCAATTGGGATGTCTTTACACAGCGCTGGCAA 135  
136 CGCAGCAACGGCCGCAACGGCGATGTGGCTTGACATCACTACCAACGCTACGATGAGAC 195  
136 ACCCTGACGGTGACACAGAGACGTTGCGTGACATTATACCGATACAGGAAGAT 195  
196 TTGTATCTCTTGACCAAGTACGGCGCAAGGCTTCCCTTCTGTGTGTGGCGG 255  
196 ATCCACTGATGAAAGATAGGTTAGACGCTTACAGGTTCTCTATCTCTGGCCAGA 255  
256 ATCATTCCTCCGCGCGGAGCTGATCCGCTCAACGAGAGGAATGAGTTTACAGC 315  
256 ATTATGCCAGATG-----GAAGACATCAACCAAAAGGTGTGATTTCTACAC 306  
316 AAACATGATGACCCCTGTTAGCGCGGATATCAACGCTTGGTGAATTGTACACTGG 375  
307 AGACTGTTGATGAGCTTTGAAAGATGATATCATACATTCGTAACACTATACCTGG 366  
376 GATTCGCTCAGCGCTTCAACGATCGCTATGAGGCTGCTCAACGTTGAAGAGTCCAG 435  
367 GACTTACCTTACCCACTTATGAAA--AAGTGTAGTGGCTTAAACC--AGATATGCG 420  
436 CTGAGCTTGGAGGGATGAGGATGAGTGTGCTTTGAACGTTTGGGAGCGAGTCCGAAC 495  
421 CTCTATTACAGACATACGCAACGTTATGTCAACGACTCGGTATCTGTGAACAT 480  
496 TGATCACCATCAACGAMCCCTGATTCAAGCCATCTATGATATGCCAGCGCAGCAAC 555  
481 TGATTAACACTGAACAGACATGTGTCTTTCTCGGGTTATTAACAGGAGAGCAT 540  
556 GCCCGGGG 563  
541 GCCCGGGG 548

RESULT 4  
US-09-344-510B-6  
Sequence 6, Application US/09344510B  
Patent No. 6579850  
GENERAL INFORMATION:  
APPLICANT: Nabeshima, Youichi  
Kuroo, Makoto  
Sekine, Susumu  
Iida, Akihiro

TITLE OF INVENTION: No. 6579850e1 Polypeptide, No. 6579850e1 DNA and No. 6579850e1  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States

ZIP: 10112-3801  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
COMPUTER: Compaq DeskPro EN  
OPERATING SYSTEM: Windows 98  
SOFTWARE: Wordpad  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/344, 510B  
FILING DATE: 25-Jun-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/04585  
FILING DATE: 12-DEC-1997  
APPLICATION NUMBER: JP 347871  
FILING DATE: 26-DEC-1996  
APPLICATION NUMBER: JP 205815  
FILING DATE: 31-JUL-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Peiray, Lawrence S.  
REGISTRATION NUMBER: 31865  
REFERENCE/DOCKET NUMBER: 766.32  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 218-2100  
TELEFAX: (212) 218-2200  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3163  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: human  
IMMEDIATE SOURCE:  
LIBRARY: kidney  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9...3047  
IDENTIFICATION METHOD: E  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-344-510B-6

Query Match 8.8%; Score 128.2; DB 4; Length 3163;  
Best Local Similarity 52.1%; Pred. No. 1.8e-23;  
Matches 407; Conservative 0; Mismatches 338; Indels 36; Gaps 4;

130 CCATCCGCACCAACGCGCCCAACGCGGATGTGGCTTGGATCACTACCCCTACGAT 189  
348 CCTCCCGCTGACGCGCCGACCGGAGAGTACGAGCAAGCTAACCAACGCTTTC 407  
190 GAGGACTTGTATCTTGACCAAGTACGGGCAAGAGGCTTACCGCTTCTCTGTGTGG 249  
408 CGGACACGAGGCGCTGCGGAGCTCGGGTACTACATCACTACCGCTTCTCATCTGTGG 467  
250 TCGGATCATTTCCCTCGCGGACGAGCTGATCCGCTCAACGAGAGGGAATTGATTT 309  
468 GCGGAGTGTCTCCCAATGACAGCGGGGGGTCC--CAACGCGAGGGGTGCGCTAC 524  
310 TACAGCAACTGATTGAACGCTTGTGAGGCGGGGATATCAACGCTTGGTGAATTGAT 369  
525 TACCGGCGCTGTGAGCGGCTGCGGAGCTGGGCTGCGACCGTGTGACCTGTAC 584  
370 CACTGGATCTGCTCAGAGCGCTTCAAGATCGCTATGAGGCTGGGCTCAACGTTGAAG 429  
585 CACTGGAGCTGCGCCAGCGCTGACAGAGCGCTTACGCGGCTGGGCAACCGCGCTTG 644  
430 GTCCAGCTGACTTTGAGCGGATGAGAGTGTGCTTGAACGTTTGGGGAACGAGTC 489  
645 GCGGAC--ACTTACAGGATTAAGGAGAGCTGTGCTTCGCACTTCCGCGGTACAGTTC 701  
490 CAGACTGATCACTCAACGAMCCCTGATTCAGGCCATCTATGATATGCAACCGGC 549  
702 AAGTACTGATCACTCAACGAMCCCTTACGTGTGTGCTGCGGCAACGCGCTACCGCGG 761

QY	550	AGCAACGCGCGGGGAGGAGCAGATTAAACAAGCATCTCACCGAGGGGAAACACTGCCACT	603
Db	762	CGCCTGGCGCCCCGGGATCTCGGGGAGCGCCGCGCTC-----	797
QY	610	GAGCCGTGGCTCGCTGTGAAGAGCCAGATTCATGAGCCATGCGCGCGGTGAGCTTAC	669
Db	798	---GGGTACTCGTGGGCGACAACTCTCTCTGGCTCATGCGAAAGTCTGGCATCTCTAC	854
QY	670	AGCAGGGACTTTCGCCCTCGCAAAAAGGCGCAGATCGGCATCTCGCTCAACGGCGACTAC	729
Db	855	AATCTTCTTTCGCTCCGCTCCACTCAGGGAGGTGAGGTGTCATTTGCCCTTAAGCTCTACATGG	914
QY	730	TATAGAGCCCTGGGAGACGATGAGGCTCTGGGACAAGAGGCGTGTGACAGCAAGATGGAA	789
Db	915	ATCATATCTCTGAAGATGACCGACCCACGACATCAAAAGATGTCAAAAATCTC--TGGAC	971
QY	790	TTTCACATTTGGCTGGTTGGCAATCCCATCTTCTTGAAGAAGGACTATCCAGAGCATG	849
Db	972	TTTGTACTAGTTGGTTTGGCCAAACCGTATTTATTTGATGGTGACTATCCGAGAGCATG	1031
QY	850	AAGAGCAGCTGGGCGAGAGGCTTTCAGCCCTCACTCCGCGGACTTTGCCATCTCTAAT	909
Db	1032	AAGATATACTTTCACTATCTATCTGCTGATTTTACTGAATCTGAGAAAAAGTTTCATCAA	1091
QY	910	G 910	
Db	1092	G 1092	

RESULT 5  
US-09-344-510B-7  
; Sequence 7, Application US/09344510B  
; Patent No. 6579850  
; GENERAL INFORMATION:

```

; GENERAL INFORMATION:
;
APPLICANT: Nabeshima, Youichi
;

```

Kuroo, Makoto

Tida Akihiro  
sexine, susumu

TITLE OF INVENTION: NO. 65798

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick

STREET: 30 Rockefeller P

CITY: New York

STATE: New York  
COUNTY: United States

COONKIN: 511224 511224  
ZIP: 10112-3801

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 1

OPERATING SYSTEM: Windows

SOFTWARE: Wordpad

CURRENT APPLICATION DATA:

APPLCATION NUMBER: US/O  
FILING DATE: 25-Jun-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/

APPL. NO. 12-DEC-1997

FILING DATE: 26-DEC-1996

APPLICATION NUMBER: JP 2

FILING DATE: 31-JUL-1997

NAME: PERRY, Lawrence S.

REGISTRATION NUMBER: 318

REFERENCE/DOCKET NUMBER:

TELEPHONE: (212) 218-2100  
; INFORMATION  
; TELECOMMUNICATIONS  
; :

TELEFAX: (212) 218-2200

; INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3435

TYPE: nucleic acid

```

? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? ORIGINAL SOURCE:
? ORGANISM: human
? IMMEDIATE SOURCE:
? LIBRARY: kidney
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 9..1655
? IDENTIFICATION METHOD: E
? SEQUENCE DESCRIPTION: SEQ ID NO: 7:
? US-09-344-510B-7

```

	Query Match	Similarity	8.8%	Score 128.2	DB 4	Length 3435
	Best Local	Similarity	52.1%	Prod. 1.8e-23		
	Matches	Conservative	0	Mismatches 338	Indels 36	Gaps 4
Qy	130	CCATCGCGCAACCAAGCGCGCAACGCGGATGTGGCTTGGCATCTTACCAACCGCTACGAT	189			
Db	348	CCGTCGCGCGCTGCAGCGCCCGCCACCGGGGACGTAGCCAGCCACAGCTTCAACCAACGCTTTC	407			
Qy	190	GAGGACTTTGATCTCTTTGACCAAGTACGCGCAAAAGGCTTACCGCTTCTCTTGTCTGG	249			
Db	408	CGGCAACAGGAGGCGCTGGCGGAGCTCGGGGTACTCACTACCGCTTCTCATCTGCTGG	467			
Qy	250	TCCGCGATCTATCCCTCTCGGCGGCGAGGCTGGATCCGCTCAACGAGAGGGAATTGACTTT	309			

Accession	Sequence	Position
Dp	GCGGAGTGTCTCCCAATGGCAGGCGGGGCTCCC---CAACGCGGAGGGGCTGCCTAC	524
Dp	468	
Qy	TACAGCAAACTGATTTGACGCGCTCTTTGAGGCGGGGATATCA GCGCTTGGGGATCTTTGTAC	369
Dp	525 TACGCGCGCTCTGTGGAGCGGCGCTGCGGAACCTGGCGCTGACGCCGTGGTCACTCTGTAC	584
Qy	370 CACTGGGATCTGCTCTCAGCGCTTCAAGATCGCTATGAGAGCTGCTCAA GCTTGAAGAG	429
Dp	585 CACTGGGACCTGCGCCCAAGCGCTGTGCAAGACGCTTACGCGGCTGGGCCAACC GCGGCCCTG	644
Qy	430 GTCCAGCTGGAACCTTTAGACGGGTATGCAAGGTGTGCTTTAAAGCTTTTGGGGACCAAGTC	489

[illegible]

Accession	Sequence	Position
Qy	AGCAGGACATTTGGCCCTCGCAAAAAGGCGCAATTCGGATCTTCGCTCAACGGAGATAC	729
Db	AATACCTCTTCGTCGCCATCTCAGGAGAGTCAGGTGCCATTCGCCCTAACTCTCAATCG	914
Qy	TATGAGCCCTGGGACAGCATGAGCCTCGGGACAAAGAGGCTGCTGAGCAAGCATGGAA	789
Db	ATCATCTCTCGAAGATGAGCCGACACACAGATCAAGAAATGTCMAAAATTC---TGGAC	971
Qy	TTTCACATTTGGCTGTGTTGCCAATCCCATCTTCTTGAAGAAGACTATCCAGAGCATG	849
Db	TTTGTACTAGTGTTGGTTTGCCAAACCCGTAATTTATTGATGGTACTATATCCGAGAGCATG	1031

Db 1032 AAGATAACCTTCATCTATTCTGCCGATTTACTGATCTGAGAAAAAGTTCATCAA 1091  
QY 910 G 910  
Db 1092 G 1092





Db	482	ATATAGTAAACCAACAATTGCTTCTTCTTCATGCTGCTGTGAAACATATACAAGAGA	541
Oy	677	ACTTTGCGCCCTCCGAAAGGGCCAGATGGGCACTTCGCTCAACGGCGACTACTATGAGC	736
Db	542	AATATCAGGCAATATCAAAAGGGGCGAGATAGGGATTAACCTAGTACTTATTTGATGATTC	601
Oy	737	CCTGGGACAGCAATAGAGCTCTGGGACAAAGAGAGCTCTGAGCGACGATGAATTTACA	796
Db	602	CCT----ACTCCATTCGAAAGCGCGACAAAGATGCGACACACAGAGCCCTTGATTTCATGT	658
Oy	797	TTGCGCTGTTTGCCAAATCCCATCTTCTTGAAGAGACTATTCAGAGAGCATGAAGAAGC	856
Db	659	ATGATGTTGTTTATTTAGAGCCATTAAAGCTT---TGCTCAATATCCAAAAAGCATCGTAGAC	715
Oy	857	AGCTGGCGAGAGGGCTTCCAGCCCTCACTCCGCGGGA	893
Db	716	TCGTTGGTAAAGATTACCAAGGTTCACTTAAGAACA	752

```

RESULT 9
US-09-122-230-8
: Sequence 8: Application US/09122230A
: Patent No. 5973228
: GENERAL INFORMATION:
: APPLICANT: Carlson, et al.
: TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin
: TITLE OF INVENTION: Composition in Plants
: FILE REFERENCE: 50532
: CURRENT APPLICATION NUMBER: US/09/122,230A
: EARLIER APPLICATION NUMBER: U.S. 60/053,566
: EARLIER FILING DATE: 1997-07-24
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1542
: TYPE: DNA
: ORGANISM: Pinus contorta
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1542)
: US-09-122-230-8

```

Query Match	8.6%	Score 124.8	DB 2	Length 1542
Best Local Similarity	54.4%	Pred. No. 1e-22		
Matches 299	Conservative 0	Mismatches 242	Indels 9	Gaps 2
Qy	20	TGCCCAAGACATTGATGAGGCTTCGCAACGCGCCCTACAGATCGAAGCGCGCTCA	79	
Db	89	TCCCCTCAGATTTCAATGTTCCGCAAGCCTCTTACGCGTATAGATGAAAGAGCAGTCC	148	
Qy	80	AAGAAGTGCGCGCGCGCCCTGCATCTGAGACAGCTACTGCCACTGAGCCATCGCGCA	139	
Db	149	GAGAAAGATGGCAAGGATCCTAGCACAATGGAGCGCTTAACATATATGCTGTAGATTA	208	
Qy	140	CCAAAGCGGCCCAAGCGCGATGTGCTTGGGATCACTACACCGCTATGAGACTTTGG	199	
Db	209	AAGATAGAGCAGATGAGACGTTGCGACGATCAACCAATATCACGATATATGGAAGATATCG	268	
Qy	200	ATCTCTTGACCAAGTACGAGCGCAAGGCGCTACCGCTTCCTCGTGTGAGTGGCGGATCA	259	
Db	269	AGCTTATGGCTTCACTTGGACTAGATGCGCTATAGATTTCTCCATATCTCGTCTCGAATCC	328	
Qy	260	TTCCCTTCGCGGCGAGCGCTGATCCCGTCAACGAGAGGGAATGATTTTACAGCAAC	319	
Db	329	TTCCAGAGGAAG-----AGGTGAATTTACATATGCTGTGGGATTTGAATTTACATATTC	382	
Qy	320	TGATTTGACGCGCTGTTGAGCGCGGATTCACGCTTTGGGTGACTTTGTAACAATCGGATC	379	
Db	383	TGATTTGACGCTTCTTCGCAAAATGAGGATTCAGCCCTGTGACATTTGTTCCATTTTCGATC	442	
Qy	380	TGCCTCAGGCGCTTACGATCGCTATGAGGCTGAGCTCAAGTGGAAAGGTCAGCTGG	439	

Accession	Sequence	Position
Db	TTCCAAACACCTTGAAAGCTCCTATGGGGAGATGGCTGAGTCTCAATAATTAA	499
Db	443	499
Qy	ACTTTTGAACCGGATATCGAGGTTTGCTTTGAAACGTTTTGGGGACCGAGTCCAGAACTGGA	499
Db	ACTTCGAAACCCATATGAGAGATTTTGCTTCGCGGACATTCGGTGACCCGTGCAAAATATTGGG	559
Qy	TCACCATCAACGAMCCCTGGATTGAGGCGATCTATGATATAGCAGCGGACGCAACGCC	559
Db	CGACAGTGAACGAGCCAAATCTGTTTGTGCCGTGGATACCCGTGGAAATATTTCAC	619
Qy	CGGCGAGAG	569
Db	CGACGAGTG	629

```

RESULT 10
US-09-122-230-6
: Sequence 6, Application US/09122230A
: Patent No. 5973228
: GENERAL INFORMATION:
: APPLICANT: Carlson, et al.
: TITLE OF INVENTION: Confiterin Beta Glucosidase cDNA for Modifying Lignin
: TITLE OF INVENTION: Composition in Plants
: FILE REFERENCE: 50532
: CURRENT APPLICATION NUMBER: US/09/122,230A
: CURRENT FILING DATE: 1998-07-23
: EARLIER APPLICATION NUMBER: U.S. 60/053,566
: EARLIER FILING DATE: 1997-07-24
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 6
: LENGTH: 1909
: TYPE: DNA
: ORGANISM: Pinus contorta
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (183)..(1724)
: US-09-122-230-6

```

Query Match	Best Local Similarity	8.6%; Score 124.8; DB 2;	Length 1909;
Matches 299;	Conservative	0; Mismatches 242;	Indels 9; Gaps 2.
Qy	20	TGCCCAACGACTTGTGAATGGGGCTTTCGCAACGCGCGCTTACAGATCGAAGCGCCGTCA	79
Db	271	TCCCCTCAGATTTCATGTTCGGCGACAGCCTCTTCAGGGTATCAATGAAAGACAGATGCC	330
Qy	80	AAGAAAGTGGCCGCCGCCGTCCATCTGGGACACGTACTGCCACCTTGAGCCATTCGCCCA	139
Db	331	GAGAAAGATGGCAAGGGGTCTCTGACACATGGGCGCTTTAACACATATGCTCGTGAATATA	390
Qy	140	CCAAACGGCGCCCAACGGGGATGTGGCTTGGCATCTACTACACCGGTATCGATAGAGACTTGTG	199
Db	391	AAGTATGACGACATAGGAGACCTGGCAGTCGACCAATATACAGATATATATGGAAGATATCG	450
Qy	200	ATCTCTTGACCAAGTACGCGCGCAAAAGSCCTTACCGCTTCTCCCTTGTGCTGTTCGGGATCA	259
Db	451	AGCTTATGGGCTTCACTTGGACTATAGTGCCTATATATTCTTCATATCCGTGCTCGAATCC	510
Qy	260	TTCCCTTCGGCGGACGCGTGGATCCCGTCAACGAGAGAGGAATTGAGTTTATACAGCAAAAC	319
Db	511	TTCCAGAAAGGAAAG-----AGGTGMAATTAACAATGCGCGGATGGAATATTATACATATATC	564
Qy	320	TGATTGACGCCCTGTGTAGAGCGGGGTATACAGCCTTGGGTACTTTGTATACACATCGGAGATC	379
Db	565	TGATTTAGCGCTTCTTCTGCAAAATGGGATTCAGCCGTTCTGTATCAATTGTTCCATTTTCGATC	624
Qy	380	TGCCTCAGGCGCTTACAGATCGCTATATGAGAGGCTTGCTCAACGTGAAAGAGTCCAGCTGG	439
Db	625	TTCCCAAGACCTTGAAAGACTCTTATGGGGGATGGCTGAGTGCCTCAATATATTA-----CG	681
Qy	440	ACTTTGACGGTATACGAGGTTGTGCTTTGAACTTTTGGGGACCGAGTCCAGAACTCGA	499

Db 662 ACTTGGAAGCCTATGACAGATTTGCTTCGCGGCAATCGGTGACCGGTGTCAAATATTTGGG 741  
QY 500 TCACCATCAACGAMCCCTGATTTCAAGGCATCTATGATATGACCGGACGCAACGCC 559  
Db 742 CGACAGTAGACGACGCAAAATCTGTTGTGCGCTTGGATACACCGTGGAAATTTTCCAC 801  
QY 560 CGGCGAGAG 569  
Db 802 CGACGAGGTG 811

RESULT 11  
US-08-688-988-9  
; Sequence 9, Application US/08688988B  
; Patent No. 6096545  
; GENERAL INFORMATION:  
; APPLICANT: Lefebvre, Daniel D.  
; APPLICANT: Malboobi, Mohammad A.  
; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS  
; FILE REFERENCE: PPL96-03  
; CURRENT APPLICATION NUMBER: US/08/688,988B  
; CURRENT FILING DATE: 1996-07-31  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FASTSEQ for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 1829  
; TYPE: DNA  
; ORGANISM: Arabidopsis Thaliana  
US-08-688-988-9

Query Match 8.2%; Score 118.6; DB 3; Length 1829;  
Best Local Similarity 53.6%; Pred. No. 4.3e-21;  
Matches 294; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

QY 18 TCTGCCCAACGACCTTGAATGGGCTTGCAACGCGCCCTACAGATCGAAG---CGC 74  
Db 159 TTTCCCTGAAGGTTTTTTATTTGGACGCGTACTGCGGCATACCAAGGATTCGA 218  
QY 75 CGTCMAAGAGGTGCGCGCGCCGCTCCATCTGGGACACGTACTGCCACTGGAGCATC 134  
Db 219 TTTATGAAACTTGTGTGGACCGCTTATGGGACATCTACTGTGAATATTCAGA 278  
QY 135 GCGACCAACGCGCCCAACGCGCATGTGGCTTGGGACTCAACACGCGTACGATGAGA 194  
Db 279 GAGGTGCAATACGATACGCGCATGTGGCGGTGATTTCTTCATGCTTATAGAGA 338  
QY 195 CTTTGATCTCTTGACCAAGTACGGGCGCAAGCGCTTCCTTGTCGTGTGCG 254  
Db 339 TATCCACTATGAAAGATCTAAACACAGACGCTTTGAATGTCTATCGCATGGCCA 398  
QY 255 GATCATTCCTCGCGCGGAGGCTGTATCCGTCACGAGAGGAAATGATTTTACG 314  
Db 399 AATATTTCTCATGGGAGAAAGAGAAAGAGTGAATGATGATGCAATTTTCA 458  
QY 315 CAATCTATTTAGCCCTGTTGAGCGGGGTATACGCGCTTGGGTGACTTTGAACCTG 374  
Db 459 CGACTCATGACGACGCTCATAAAAAATGTATTAATCTCATCTGTTTCTGACTG 518  
QY 375 GGATCTGCTCAGGCGCTTCAGATCGCTATGAGAGCGGCTCAACGTGGAAGGCTCA 434  
Db 519 GGAACCTCCACAAATTTAGAAATGATATGAGGCGCTTTTAAAGC---GAAAGATTTGT 575  
QY 435 GCTGGACTTTGAGCGGTATGCGAGGTTGTGCTTTGAACGTTTGGGAGCGAGTCCGAA 494  
Db 576 GAAGAGATTTCCGAGAGTATGAGATTTTGTTCGAAAGATAGCGGTGGAAGAGTCA 635  
QY 495 CTGGATACCATCAACGAMCCCTGATTTCAAGGCATCTATGATATGCAACGCGCAGCA 554  
Db 636 TTGGATCACTTTCAATAGGACATGGGTTTCTCGACGCTGGCTATGACGTAGGCAAAA 695  
QY 555 CGCCCCGGG 563  
Db 696 GGCACCTGG 704

RESULT 12  
US-08-400-275-17  
; Sequence 17, Application US/08400275  
; Patent No. 5668295  
; GENERAL INFORMATION:  
; APPLICANT: Mahab, Samir Z.  
; APPLICANT: Malik, Vedpal S.  
; TITLE OF INVENTION: PUTRESCINE N-METHYLTRANSFERASE  
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING PUTRESCINE  
; TITLE OF INVENTION: N-METHYLTRANSFERASE, AND TRANSGENIC TOBACCO PLANTS WITH  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Neave  
; STREET: 1251 Ave. of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/400,275  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/076,681  
; FILING DATE:  
; APPLICATION NUMBER: US 07/613,160  
; FILING DATE: 14-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Creason, Gary L.  
; REGISTRATION NUMBER: 34,310  
; REFERENCE/DOCKET NUMBER: PM-1696  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1545 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-400-275-17

Query Match 7.7%; Score 111.4; DB 1; Length 1545;  
Best Local Similarity 49.9%; Pred. No. 2.9e-19;  
Matches 364; Conservative 0; Mismatches 356; Indels 9; Gaps 3;

QY 14 TAGCTCTGCCCAACGACCTTGAATGGGCGCTTGCAACGCGCCCTACAGATCGAAGCG 73  
Db 2 TGAGCTTTTCTTCTATATTTCTTGTGGGACAGCCTTTCATATTTACAGATGAAAGAG 61  
QY 74 CCGTCMAAGAGGTGCGCGCGCCGCTCCATCTGGGACACGTACTGCCACTGGAGCAT 133  
Db 62 CTTTCTCTAGTATGGGAAAGGCTTCAGCACTGGGAGCTTTTAAACCATGAGCTGTGTC 121  
QY 134 CGGCAACCAACGCGCAACGCGCATGTGGCTTGGGATCACTCAACGCGTACGATGAG 193  
Db 122 ATGTTAAGATGAAGACATGAGATGTGGCTGTGATCATAACCATGTTATTTGAGG 181  
QY 194 ACTTGATCTCTTGACCAAGTACGGCGCAAGGCGCTTACCGGCTTCTTGTCGTGTGCG 253  
Db 182 AATCAAACTCATGGCAGATATGGGTGTGAATAGCTTTGTTCTCTATCTCATGGGC-- 239



OY	254	GGATCATTTCCCTCGGCGGCGAGGCTGGATCCGTCACAGAGAGAGGAAATTGATTTTAC	313
Db	240	-AAGATTCTCGCCCAAGGSAATATTTTGGAGAGTTATATATGCGCGAATTTGACCTAC	298
OY	314	GCAAACTGATTTGACCCCTGTTGAGAGCGGAGTATCAAGCCTTGGGAGTACTTTGTACACT	373
Db	299	GTAACCTCATTTGATGTACACTCTTACAGAAAGGATCCAGCCGTTTGTCACTTAACACATT	358
OY	374	GGGATCTGCTTCAGGCGCTTTCAGATCGCTATGAGAGCTGGCTCAACGTGGAAGAGTCC	433
Db	359	TTGACATATCCACAAGAACTTGAGAGACAGATATGTGTTGGCTAA--GTTCACAGATTAC	415
OY	434	AGCTGAGACTTTGAGGCGGATATGCGAGAGTTGTCTTTGAACGTTTGGGGACCGAGTCAGA	493
Db	416	GGGATGATTTTCACTATTTCGCAACATATGCTTCAAAATCACTTGGGAGATAGAGTTAAAT	475
OY	494	ACTGGATACCATCAACGAGGCCCTGGATTCAAGGCCATCTATGGATATGCAACCGGAGACA	553
Db	476	ACTGGGTAAACGATGATGATGAGGCTTAACTTGTGGCCATTATGTGGCTATATAGATGGACCTT	535
OY	554	ACGCCCCGGGCGAGAGC--AGCATTTAACAAACACTCCACCGAGGGCAACACTGCCACTG	610
Db	536	GCCCTCCAACTCGATGCTCTGTGTTATTTTGGGAATGTAGTGTGGGGATTCAAGAAAGGG	595
OY	611	AGCGCTGCTCGCTGGAAGGCCCAAGATCATGAGCCATGCCCGCGCTGGCCGCTTACA	670
Db	596	AACCTTCATTTGACACTCAATATATGATCTCTATCTCATGCAAGATGCTGTACAGATTTACC	655
OY	671	GCAGGAGACTTTGCGCCCTTCGCAAAAAGGGCAAGTCGGCATCTTGCCTCAACGGGACTACT	730
Db	656	GCACCAAGATATCAGAAAGAGTCAAGAGAGGATATGGCACTTACTATGGGTTTCGAATGGT	715
OY	731	ATGAGCCCT 739	
Db	716	ATGAACCGT 724	

```

RESULT 13
US-09-134-000C-2890
; Sequence 2890, Application US/09134000C
; Patent No. 6617156
;
GENERAL INFORMATION:
;
APPLICANT: Lynn Doucette-Stamm et al
;
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;
FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
;
FILE REFERENCE: 032796-032
;
CURRENT APPLICATION NUMBER: US/09/134,000C
;
CURRENT FILING DATE: 1998-08-13
;
PRIOR APPLICATION NUMBER: US 60/055,778
;
PRIOR FILING DATE: 1997-08-15
;
NUMBER OF SEQ ID NOS: 6812
;
SOFTWARE: PatentIn version 3.1
;
SEQ ID NO 2890
;
LENGTH: 1458
;
TYPE: DNA
;
ORGANISM: Enterococcus faecalis
;
US-09-134-000C-2890

```

	Query Match	7.2%	Score 10.48	DB 4	length 1458
	Best Local Similarity	53.4%	Pred. No. 1.4e-17		
	Matches	267	Conservative	1	Mismatches 223
				Indels	9
				Gaps	2
QY	15 AGCTTCGCCAAGCATTTGATGAGGGCTTCGCAACGGCCGCTTACCAATGAAAGCGC	74			
Db	45 AGCATTTCCGAAAGATTTTATATGGGCTCTGCTTCGGAGCTTATCAAGTGAAAGGTGC	104			
QY	75 CGTCAAGAAAGTGGCCCGCGCCGTCATCTTGGGACACGTACTGCCACCTTGAGCCATC	134			
Db	105 TTGGATATGAAGCGGCAAGGGCGCATCTGTTGGATGATATTTGTTCCGCTCCAGAAA	164			
QY	135 GCGCACCAACGGCGCCCAACGGGATGATGGCTTGGCATCACTCCACCGCTACGATGAGAGA	194			
Db	165 AACGTTCAAGAGACAACCTGGGATTTTACCGCTGCATCATATATCAATGTTTCAAGAAGA	224			

Oy	195	CTTTGATCTCTGAGCCAGTAGTACGGGGCAAAAGGCGTACCGCTCTCTGTCGATGGCGGG	254
Db	225	TGTGGCTTTAATGAAACAACAAGTTTAAAGCATATGTTTCTGATGTCAATGGACACG	284
Oy	255	GATCATTTCCCTCGGGCGGACGAGCTGATCCCGTCAACGAGAGGAAATTGATTTTACAG	314
Db	285	AATTTTAC-----TGAAGGCGGTGTGAGTAAATCAAGGGGACATAATTTTATTC	338
Oy	315	CAAACTGATTAGCGCCCTGTGAGGCGGGGATACAGCCTTGGGTGCATTTGTACCATG	374
Db	339	AGATTTAATTGATGATGATTATTAGCGGACGGAATTGAACCAATGTAACTTTGATCATTG	398
Oy	375	GGATCTGCTCAAGCGGCTTCAAGATCGCTATGAGAGCTGGCTCAACGTGGAAGAGGTCCA	434
Db	399	GGATTTGGCACTGTTTGGCAAAAAGAAATATGTGTGGCTGG---GAATCAAGAAAAATTAT	455
Oy	435	GCTGACCTTGAGCGGCTATGCGAGGTGTGCTTTGAACGTTTGGGGACCGAGTCCAGAA	494
Db	456	TGATGATTTTGCGGCTATGCGAAAATTTTATTGACGTTTCGTTGCGCAAGATTCGCTA	515
Oy	495	CTGATCAACATCAACGAMC	514
Db	516	TTGGATTGTTAAATGAAC	535

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 29, 2004, 00:58:24 ; Search time 511.319 Seconds  
(without alignments)  
10594.130 Million cell updates/sec

Title: US-10-026-140-3

Perfect score: 1455  
Sequence: 1 atgcccagctcgtcagctctc.....gggttaagtgccgcatcaaa 1455

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PC7\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PC7US\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1454.2	99.9	1455	US-10-026-140-3	Sequence 3, Appli
2	1454.2	99.9	1455	US-10-026-140-3	Sequence 1, Appli
3	317.8	21.8	1470	US-10-369-493-37092	Sequence 37092, A
4	233.4	16.0	1688	US-10-425-114-4348	Sequence 4348, Ap
5	230.4	15.8	1794	US-10-425-114-19747	Sequence 19747, A
6	227.4	15.6	1781	US-10-425-114-30939	Sequence 30939, A
7	227.4	15.6	1931	US-10-425-114-30835	Sequence 30835, A
8	227.4	15.6	1932	US-10-425-114-33175	Sequence 33175, A
9	219.2	15.1	1737	US-10-425-114-25892	Sequence 25892, A
10	187.6	12.9	1759	US-10-425-114-13710	Sequence 13710, A
11	184.4	12.7	1735	US-10-425-114-24850	Sequence 24850, A
12	184.4	12.7	1757	US-10-425-114-27741	Sequence 27741, A
13	183.2	12.6	1338	US-10-369-493-31938	Sequence 31938, A
14	182.8	12.6	1734	US-09-938-842A-1282	Sequence 1282, Ap
15	182.8	12.6	1734	US-09-938-842A-1282	Sequence 1282, Ap

16	180.8	12.4	1472	US-10-369-493-31813	Sequence 31813, A
17	180.2	12.4	1735	US-10-425-114-8342	Sequence 8342, Ap
18	180.2	12.4	1829	US-10-424-599-26468	Sequence 26468, A
19	178.2	12.2	2016	US-10-147-493-253	Sequence 253, App
20	178.2	12.2	2016	US-10-145-127-253	Sequence 253, App
21	178.2	12.2	2016	US-10-160-503-253	Sequence 253, App
22	178.2	12.2	2016	US-10-143-118-253	Sequence 253, App
23	178.2	12.2	2016	US-10-144-993-253	Sequence 253, App
24	178.2	12.2	2016	US-10-158-787-253	Sequence 253, App
25	178.2	12.2	2016	US-10-081-056-347	Sequence 347, App
26	178.2	12.2	2016	US-10-028-072-253	Sequence 253, App
27	178.2	12.2	2016	US-10-121-049-253	Sequence 253, App
28	178.2	12.2	2016	US-10-123-904-253	Sequence 253, App
29	178.2	12.2	2016	US-10-140-470-253	Sequence 253, App
30	178.2	12.2	2016	US-10-175-746-253	Sequence 253, App
31	178.2	12.2	2016	US-10-176-918-253	Sequence 253, App
32	178.2	12.2	2016	US-10-176-921-253	Sequence 253, App
33	178.2	12.2	2016	US-10-137-865-253	Sequence 253, App
34	178.2	12.2	2016	US-10-140-474-253	Sequence 253, App
35	178.2	12.2	2016	US-10-142-431-253	Sequence 253, App
36	178.2	12.2	2016	US-10-143-114-253	Sequence 253, App
37	178.2	12.2	2016	US-10-140-002-253	Sequence 253, App
38	178.2	12.2	2016	US-10-142-419-253	Sequence 253, App
39	178.2	12.2	2016	US-10-123-262-253	Sequence 253, App
40	178.2	12.2	2016	US-10-142-423-253	Sequence 253, App
41	178.2	12.2	2016	US-10-121-050-253	Sequence 253, App
42	178.2	12.2	2016	US-10-141-755-253	Sequence 253, App
43	178.2	12.2	2016	US-10-143-032-253	Sequence 253, App
44	178.2	12.2	2016	US-10-123-108-253	Sequence 253, App
45	178.2	12.2	2016	US-10-123-236-253	Sequence 253, App

#### ALIGNMENTS

RESULT 1  
US-10-026-140-3  
; Sequence 3, Application US/10026140  
; Publication No. US20030114330A1  
GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegeburt, Frits  
; APPLICANT: Ward, Michael  
; APPLICANT: Yao, Jian  
TITLE OF INVENTION: BG5 Beta-Glucosidase and Nucleic Acids  
FILE REFERENCE: Encoding the Same  
CURRENT APPLICATION NUMBER: US/10/026.140  
CURRENT FILING DATE: 2002-03-26  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 1455  
TYPE: DNA  
ORGANISM: Trichoderma reesei  
US-10-026-140-3

Query Match 99.9%; Score 1454.2; DB 14; Length 1455;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCCGAGTCGTAGCTTGCCCAAGATTGATGGGCTTGCACGCGCCTAC	60
DB	1	ATGCCGAGTCGTAGCTTGCCCAAGATTGATGGGCTTGCACGCGCCTAC	60
QY	61	CAGATGAAAGCGCGTCGAAAGAGTGCGCGCGCGCTTCATCTGGACAGCTATGC	120
DB	61	CAGATGAAAGCGCGTCGAAAGAGTGCGCGCGCGCTTCATCTGGACAGCTATGC	120
QY	121	CACCTGAGGCGATCGGCGACCAACGCGCGCGAGATGTGCTTGCATCTACTACAC	180
DB	121	CACCTGAGGCGATCGGCGACCAACGCGCGCGAGATGTGCTTGCATCTACTACAC	180

```
QY 181 CGCTACGATGAGGACTTTGATCTTGA CCAAGTACGGGCGAAGGCGCTTCTCC 240
DB 181 CGCTACGATGAGGACTTTGATCTTGA CCAAGTACGGGCGAAGGCGCTTCTCC 240
QY 241 TTGTGCTGATCGCGGATCATTTCCCTCGGCGGACAGGCTGATCCCTCAAGAGAGGGA 300
DB 241 TTGTGCTGATCGCGGATCATTTCCCTCGGCGGACAGGCTGATCCCTCAAGAGAGGGA 300
QY 301 ATTGATTTTACAGAACTGATTTGACGCGCTTGTGAGGCGGGGTATACCGCTTGGGTG 360
DB 301 ATTGATTTTACAGAACTGATTTGACGCGCTTGTGAGGCGGGGTATACCGCTTGGGTG 360
QY 361 ACTTGTACCACTGGGATCTGCTCAGAGGCTTCAAGATCGCTATGAGGCGTGGCTAAC 420
DB 361 ACTTGTACCACTGGGATCTGCTCAGAGGCTTCAAGATCGCTATGAGGCGTGGCTAAC 420
QY 421 GTGGAAGAGTTCAGCTGGACTTTGAGCGGTATGCGAGGTTGTGTTGAAAGTTTGGG 480
DB 421 GTGGAAGAGTTCAGCTGGACTTTGAGCGGTATGCGAGGTTGTGTTGAAAGTTTGGG 480
QY 481 GACCGAGTCCAGAACTGGATGACATCAACAGMCCCTGGATTCAGGCGCATCTATGATAT 540
DB 481 GACCGAGTCCAGAACTGGATGACATCAACAGMCCCTGGATTCAGGCGCATCTATGATAT 540
QY 541 GCCACCGGACCAACGCGCCCGGGGAGAGGACATTAACAAGCACTCCACGAGGCGAAC 600
DB 541 GCCACCGGACCAACGCGCCCGGGGAGAGGACATTAACAAGCACTCCACGAGGCGAAC 600
QY 601 ACTGCTACTGAGCGCTGCTGCTGGAAGAGCCCAAGATCATGAGCCATGCGCGCGTG 660
DB 601 ACTGCTACTGAGCGCTGCTGCTGGAAGAGCCCAAGATCATGAGCCATGCGCGCGTG 660
QY 661 GCCGCTTACAGCAGGAGCTTTCGCCCTCGCAAAAGGCGCGAGTGGGATTCGCTCAAC 720
DB 661 GCCGCTTACAGCAGGAGCTTTCGCCCTCGCAAAAGGCGCGAGTGGGATTCGCTCAAC 720
QY 721 GCGCACTATATGAGCCCTGGGACAGCAATGAGCTCGGGA CAAAGAGCTGTAAGCGA 780
DB 721 GCGCACTATATGAGCCCTGGGACAGCAATGAGCTCGGGA CAAAGAGCTGTAAGCGA 780
QY 781 CCGATGGAATTTCAATTTGGGTGGTTTGGCAATCCCATCTTTTGAAGAGACTATCCA 840
DB 781 CCGATGGAATTTCAATTTGGGTGGTTTGGCAATCCCATCTTTTGAAGAGACTATCCA 840
QY 841 GAGAGCATGAGAAGAGCAGCTGGGCGAGAGGCTTCAACCCCTACTCCCGCGGACTTTGGC 900
DB 841 GAGAGCATGAGAAGAGCAGCTGGGCGAGAGGCTTCAACCCCTACTCCCGCGGACTTTGGC 900
QY 901 ATCTCTCAATGCGGAGAGACCGACTTCTAGCGCATGATTA CACATCTCCAGTTGCG 960
DB 901 ATCTCTCAATGCGGAGAGACCGACTTCTAGCGCATGATTA CACATCTCCAGTTGCG 960
QY 961 CGCACCTTACAGCGGTCCGCTCCCGAGAGCGGATCATCTCGGCGCATCCAGAGCAG 1020
DB 961 CGCACCTTACAGCGGTCCGCTCCCGAGAGCGGATCATCTCGGCGCATCCAGAGCAG 1020
QY 1021 GAGAAATAGAGCGGACCGCTTGGGAGAGAGCGGCTTCGCTGCTGCGCTCTGTC 1080
DB 1021 GAGAAATAGAGCGGACCGCTTGGGAGAGAGCGGCTTCGCTGCTGCGCTCTGTC 1080
QY 1081 CCGGACATGTTTCCGAGAGCATCTGCGCGGGGTGATACGGGCTGTA CCGCAAGCCCATCTAC 1140
DB 1081 CCGGACATGTTTCCGAGAGCATCTGCGCGGGGTGATACGGGCTGTA CCGCAAGCCCATCTAC 1140
QY 1141 ATACCGAGAAACGATGATCCGCTGCGAGAGAGAAATGAGTGCAGAGGCGCTGTC 1200
DB 1141 ATACCGAGAAACGATGATCCGCTGCGAGAGAGAAATGAGTGCAGAGGCGCTGTC 1200
QY 1201 AACGACCCCTTCCGATCCGATCTTGAATCGCATCTGATGATTTCCAAAGGCGCAT 1260
DB 1201 AACGACCCCTTCCGATCCGATCTTGAATCGCATCTGATGATTTCCAAAGGCGCAT 1260
QY 1261 ACCGAGAGCGGCTGCTGCAAGGGGACTTTGCGTGGGCGTTGCTCGATTAACCTTGGAA 1320
```

```
DB 1261 ACCGAGAGCGGCTGCTGCTGCAAGGGGACTTTGCGTGGGCGTTGCTCGATTAACCTTGGAA 1320
QY 1321 TGCTCAGATGATCTACGAGACCAATTCGCGGTACAGCTTCAACAGACTACCAACCTTCAAG 1380
DB 1321 TGCTCAGATGATCTACGAGACCAATTCGCGGTACAGCTTCAACAGACTACCAACCTTCAAG 1380
QY 1381 CGCACGCCCAAGAACTTGCCTTGGTCTTCAAGACATGTTTGGCGCGCGAGAGGTT 1440
DB 1381 CGCACGCCCAAGAACTTGCCTTGGTCTTCAAGACATGTTTGGCGCGCGAGAGGTT 1440
QY 1441 AAAGTGGCGGCATPA 1455
DB 1441 AAAGTGGCGGCATPA 1455

RESULT 2
US-10-026-140-1
; Sequence 1, Application US/10026140
; Publication NO. US20030114330A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGLs Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ. ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO. 1
; LENGTH: 1991
; TYPE: DNA
; ORGANISM: Trichoderma reesei
; US-10-026-140-1

Query Match 99.9%; Score 1454.2; DB 14; Length 1991;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTCGCTACCTGCGCCCAACGACTTTGAATGGGCTTCGCAACGCGCGCTAC 60
DB 62 ATGCCGAGTCGCTACCTGCGCCCAACGACTTTGAATGGGCTTCGCAACGCGCGCTAC 121
QY 61 CAGATGAAGGCGCGCTCAAGAGAGTGGCCGCGCGCTTCATCTGGGACAGATCTGC 120
DB 122 CAGATGAAGGCGCGCTCAAGAGAGTGGCCGCGCGCTTCATCTGGGACAGATCTGC 181
QY 121 CACTGAGCATTCGCGGCAACAGCGCGCAACGCGGATGGCTTGGATGATACATCAC 180
DB 182 CACTGAGCATTCGCGGCAACAGCGCGCAACGCGGATGGCTTGGATGATACATCAC 241
QY 181 CGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCC 240
DB 242 CGCTACGATGAGGACTTTGATCTTTGACCAAGTACGCGCAAGGCTTACCGCTTCTCC 301
QY 241 TTGTGCTGATCGCGGATCATTTCCCTCGGCGGACAGGCTGATCCCTCAAGAGAGGGA 300
DB 302 TTGTGCTGATCGCGGATCATTTCCCTCGGCGGACAGGCTGATCCCTCAAGAGAGGGA 361
QY 301 ATTGATTTTACAGCAACTGATTTGACGCGCTGTTGAGGCGGGGTATCAAGCTTGGGTG 360
DB 362 ATTGATTTTACAGCAACTGATTTGACGCGCTGTTGAGGCGGGGTATCAAGCTTGGGTG 421
QY 361 ACTTGTACCACTGGGATCTGCTCAGAGGCTTCAAGATCGCTATGAGGCTGCTCAAC 420
DB 422 ACTTGTACCACTGGGATCTGCTCAGAGGCTTCAAGATCGCTATGAGGCTGCTCAAC 481
QY 421 GTGGAAGAGTTCAGCTGGACTTTGAGCGGTATGCGAGGTTGTGCTTGAAGCTTTGGG 480
DB 482 GTGGAAGAGTTCAGCTGGACTTTGAGCGGTATGCGAGGTTGTGCTTGAAGCTTTGGG 541
```

```

: Publication No. US20030233675A1
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xiandeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: CURRENT FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: PRIOR FILING DATE: 2002-02-21
: NUMBER OF SEQ. ID NOS: 47374
: SEQ. ID NO 37092
: LENGTH: 1470
: TYPE: DNA
: ORGANISM: Aspergillus nidulans
: US-10-369-493-37092

```

Qy	20	TGCGCAACACATTTGAATGGGGCTTCGGAACGGCGCGCTTAACAATGGAAGGCCCTCA	79
Db	2	TGGCAATACATCTTTCATGGAATATGCCACGGCGCGCGCAGAGTTGAAGAACGTGGA	61
Qy	80	AMAGAGTGGCGCGGCCCGCTTCATCTGGGAACGACTCCACCTGGAGCCATCGGCA	135
Db	62	ACAAAGATGGCAAGGGTCCCTTCGATCTGGGACACATTCGGGCATATCCAGGCAAAAGTCA	121
Qy	140	CCAAAGCGCCCAACGGCGCATGTGGCTTGCGATCACTACACACGGTACGATAGGACTTTC	195
Db	122	AMACAAATAGCAATGGCATGACGCTGTGAGGTTCTACGACTTTTACCGGAAAGATGTTG	181
Qy	200	ATCTCTTGACCAAGTACGGCGCAAGGCGCTACCGCTTCCTTGTGTTGTCGGCATCA	255
Db	182	CCGTATGAAAGTCGATGGGTGTCACACGGCTTATCGATTCGCTTCCGTTCCTGGTCGCGCATAA	241
Qy	260	TTCCCCCTGGCGCGCAGGCTGATATCCCGTCAACGAGAGGGAAATGAGTTTTCACGCAAC	319
Db	242	TCCCACTTCGGAAGCGCCGACGACCGCGTCAATGACAAAGGATCAAGTACTACGAGACC	301
Qy	320	TGATTTGACCCCTGTTTGAAGCGGGGTATCACGCCCTTGGGTGACTTTGTAACACTGGATC	379
Db	302	TGGTAAAGAGGTACTACATTAATGGGAATTAGGCATTTGTAAACCTCTTTCACGTGGATG	361
Qy	380	TGCTTAGGCGCTTCAACGATCGCTATGGAAGCTGGCTCAACGTGGAAGAGTCCAAGTTG	439
Db	362	TTCCGAGGCTTTGAGAGATCGGTACGGGGCATGTGTAACAAAGAAAGTTCAATTCCTG	421
Qy	440	ACTTTGAGGGATATGAGAGTGTGCTTTTGAACGTTTGGGGACCGAGTCCAGAACTGGA	499
Db	422	ACTTTTCCGCTACCGCGGTCTCTTTTTCGAACGCTTAGGGSCAAAGTCCGTACATGGA	481
Qy	500	TCAACATCAAGAACCTTGGAATTACAGGCATCTATGATATGCCACCGGACAAACGCC	555
Db	482	TTTCAATTCATATAACTGGTGTGTCTCGCTGGCAGGTTATGCGGACAGTGTTCATCTTC	541
Qy	560	CGGGCAGGAGCAAGATTAAACAAGCTCAACCGAGGGCAACATCGACACTGAGGCGTGGC	619
Db	542	CTGACAGGTATCGTTTGCGCAATCTCAATGAAGAGGTGATTTCTCGACAGAACGTTCA	601
Qy	620	TGCGTGAAGGCCAGATCATGAGCCATGCCCCGGCGCGTGGCCGTCTACAGCAGGAGCT	679
Db	602	TGTCGGGCACAAAACTGTGACCCACGGCCATGTGTTCAAGCTTTACGAAAGGTGT	661
Qy	680	TTTCGCCCTTCGAAAGGGCCAGATCCGCACTCTGCTCAACGGGCACTACTATGAGCCCT	739
Db	662	TCCAAACCGAGAAAGAACATTTGGCATACCTTCATAGGCACTGGTCGGAACTTT	721





```

; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30939
; LENGTH: 1781
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLB73106C01_FLI
US-10-425-114-30939

```

```

Query Match      15 6%; Score 227.4; DB 12; Length 1781;
Best Local Similarity 52.2%; Pred. No. 8,2e-57;
Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

```

```

QY 20 TCCCAACGACTTTGATGAGGCTTCGCAACGCGCCCTACAGATCGAAGGCGCCGCTCA 79
DB 230 TCCCAACGAGGCTTCGATTCGCGGACGCGACCTCGGCTACAGATCGAAGGCGCCGCT 289
QY 80 AAGAAAGTGGCGCGCGCCCTTCATCTGGGACACCTACTGCACTGAGCCATGACCGCA 139
DB 230 CCACCAACGCGCGCGCGCCCTTCATCTGGGACACCTACTGCACTGAGCCATGAGT 349
QY 140 CCAACGCGCGCGCGCGCGCGCTTCGATCACTACCAACGCGCGCGCTTCGATCACT 199
DB 350 CAGGAAATCAAAATGGAAGACGTTGAGTGATCAATACGCTACCAAGAAAGCGTGG 409
QY 200 ATCTCTTGAACCAAGTACGCGCGCGCGCGCTTCCTCTGCTGCTGCTGCGGATCA 259
DB 410 ATCTCATGAAAGTTGAACTTGAATGCTTACCGGCTTCATCTCATGCTGCTGCGGAT 469
QY 260 TTCCCTCGGCGCGCGCGCGCTTCGATCCCGTCAACGAGGAGGAAATGATTTACGCAAC 319
DB 470 TCCCGATGCGGAGGCGG-----AAAGTCAATCCAGAAAGGTGAGCGGATTAACATAT 523
QY 320 TGATGACGCGCGCTTGAAGCGCGGATACGCGCTTGGGATGACTTTGATCACTGAGATC 379
DB 524 TGATGAACTATCTGCTTGAAGGAGGATGAGGCTTGAAGCGGAAAGTGGCGGACTTG- 583
QY 380 TGCCCTCAGGCGCTTCACATGCTATGAGGCTGCGCTCAACGTTGAAAGGCTCCAGCTGG 439
DB 584 TTCCCTTTCGCGCTTGAAGAAATATGAGGCTGTTAAGCGGAAAGTGGCGGACTTG- 642
QY 440 ACTTGACGCGGATGCGAGGTTGCTTGAACGTTTGGGAGACCGAGTCCAGAACTGGA 499
DB 643 --TTTACGACTATGCTGCTGCTTCTTTTAAAGCTTACGCGGATGCGGTAAGGACGCTG 700
QY 500 TCACCATCAACGAMCCTGATTCAGGCGCATCTATGATATGCGAGCGGAGCAACGCC 559
DB 701 TTACATTCATGAGCAAGGATAGAGGCTACTTGGCTATGACAGGCGGCAAAATCTTC 760
QY 560 CGGCGAGAGAGCATTTAACAAGACCTTCACCGAGGCGCAACACTGCACTGAGCGCTGGC 619
DB 761 CTCAAAAGGTG-----ACGAGATGCGCTGCTGTTGGAATTCAGCAACGAACTTACA 814
QY 620 TCGCTGGAAGGCGCGAGATCATGAGCCATGCGCGCGCGCTGCTACAGAGGAGACT 679
DB 815 TAGTTGCTCAATATTTTCTTTGGCAATGCTACTGAGATTTGCAAGATCCGATCGAAT 874
QY 680 TTGCGCCCTCGCAAAAGGCGCGAGTGGCATCTGCTCAACGCGGACTACTATGAGCCCT 739
DB 875 ATCAGGCTGCTCAAGAGGTTAAGTGGAAATGCTGCTGAGACTTCACTGGATAGAGGCTC 934

```

```

QY 740 GGAACAGCATGAGCTTCGCGAGAGGAGGCTCTGAGCGACGAGTGAATTTACATTTG 799
DB 935 T---TACAACTCAGCTGATGAGCAAGACAGAGCCCAAGAGCGAGGACTTCCACTTTG 991
QY 800 GCTGTTTGCCAAATCCCATCTTCTTGAAGAAAGACTTTCAGAGAGGATGAAAGAGAGC 859
DB 992 GCTGTTTGTGATCC---ATTGATTAACGAGCACTATTCACAGATTAATGCAAGATTCG 1048
QY 860 TGGCGAGAGGCTTCAGACCTCAGCTCCGCGAGACTTGGCATCTTCATATGCGGAGAGA 919
DB 1049 TGAAGAGAGGCTCCAGGTTCACTCTAGAGAGGTAACATGCTTAAGAGGCTCGCAG 1108
QY 920 CCGACTTCTAGAGGATTAATTC---TACATCTCCAGTTTGGCGCCACCTAGACGCTC 976
DB 1109 ACTACATCGGATATCAAGAGTATACATTCAGCTACATGAAAGGCGAGAACTGCTCAGC 1168
QY 977 CCGTCCCGGAGAGGAGCTATCTCGGCGCATCCATGAGCAACGAGAAATTAAGAGCGCA 1036
DB 1169 TGGGCGCCAGTACTGCTGCGATTTGGCAGGTTCAATATGTTTTCAGCGCAATGGCA 1228
QY 1037 GCGCGTTGCGAGAGAGAGGCGGCTCGCTGCGCTTCCTGCGCGAGATGTTCCGGA 1096
DB 1229 AACGATTTGAGACCAAGCGAATTTAAGTGGCTTACATTCGCCCGAGCGGAGATGAG 1288
QY 1097 AGCATCTCGCGCGGCTGATAGGCTGTAGCGCTGAGCGAAGC---CATCTATCATCCGAGAA 1153
DB 1289 GTGCGTGAATCTTCAAGAGAGAGATGAGGAAATCAAGATCTACATTAACGAGAAAG 1348
QY 1154 GATGCCCGTCCCTGAGAGAGAGAAATACATGACGAGAGAGGCGCTCAACGAGCCCTTCC 1213
DB 1349 GAATGAGCAAGCTTGA-----AATTGACCGGAAACAGATCTGCGCGAGCGACGA 1402
QY 1214 GCATCCGATCTTGAAGTCTGAGCTGAGCTGATTCGATTTCCAGGCGATTAACCGAGAGCG 1273
DB 1403 GGGTGGGCTTCTACAGAGGCTATCATGCGGAGCTGAAAGGCGCATAGACAG---GGAG 1459
QY 1274 TCGTTCGAAAGGAGTCTTTCGCTGAGGCGTGTGCTGATTAATGGAATGCTAGATGCT 1333
DB 1460 CGAAGCTGAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1519
QY 1334 ACGAGCCAGATTTGCGGCTGACGTTCAAGACTTACACCACTTCAAGCGAGCGCCAGAG 1393
DB 1520 ACTGTCGCAAGTTGCGGATGCTTACGTGAGCTTCAACGCTCGAAGCGCACCGGAGG 1579
QY 1394 AGTGTCCCTGCTGCTTCAAGGACATGCTT 1422
DB 1580 CGTGCGCTTACTGCTTCAAGGACATGCTT 1608

```

```

RESULT 7
US-10-425-114-30835
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313) B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30835
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: UC-ZMFLB73086G03_FLI

```



US-10-425-114-30835

Query Match 15.6%; Score 227.4; DB 12; Length 1931;  
 Best Local Similarity 52.2%; Pred. No. 8.4e-57;  
 Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

QY 20 TGCCCAACGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATCGAAGGCCCGCTCA 79  
 Db 236 TCCCAAGGGGTTCTGTTTCGGGACGGGACGTCGGGTACAGAGTCGAGGGCGCCGCT 295  
 QY 80 AAGAAGTGGCCCGCGCCGCTTCATCTGGGACAGTACTGCGACCTGGAGCCATCGGCA 139  
 Db 296 CCAACCAACGGCCCGGGCCCTCCATCTGGGATTCATTCGCGCACGTCCTCCAGAAATATTG 355  
 QY 140 CCAACGGCGCAACGGCGATGGCTTGGCATCTACCAACCGCTACAGATGAGGACTTTG 199  
 Db 356 CAGGGAATCAAAATGAGAGACCTTGCAATGATCAATACATCGCTTACAGAGAACGTTG 415  
 QY 200 ATCTTTGACCAAGTACGCGGCAAGGCTTACCGCTTCTCTTGTCTGTGCGCGATCA 259  
 Db 416 ATCTCATGAAAGTTGAACTTTGATGCTTACCGGTTCTCAATCTCATGTGCGAGATCT 475  
 QY 260 TTCCCTCCGCGGACGCTGGATCCGCTCAACGAGAGGGAATGAGTTTTCACGCAAC 319  
 Db 476 TCCCGGATGGCGAAGG-----AAGTCAATCCAGAAAGGTGAGCGTATTAACAATATT 529  
 QY 320 TGAATGACGCGCTGTTGAGGCGGGGTATCACGCTTGGGATCTTTGATCAACGAGGATC 379  
 Db 530 TGAATAACTATCTGCTTCAAGAAAGGCAATGACTCTTACATCAACCTTTACCATATGATC 589  
 QY 380 TGCTCAGGCGCTTACGATCGCTATGAGGCTGGCTCAACGTGGAAGAGTCCAGCTGG 439  
 Db 590 TTCTCTTGCGCTTGAGAAAGAAATATGAGGGTGTATTAAGCGGAAATGCGGACTTG- 648  
 QY 440 ACTTTGACCGGTATGCAAGTGTGCTTTGAACCTTTGGGAGCGAGTCCAGAACTGGA 499  
 Db 649 --TTTACAGACTATGACTCTTCTGTTTAAAGCTTCGCGGATCGCGTAAAGCACTGCT 706  
 QY 500 TCAACATCAACGAGMCCCTGATTCAGGCACTATGATATGACCGGAGCAAGCCGCT 559  
 Db 707 TTAACATTCATGAGCCAGAGATAGTACGCTTCTGCTATGACAGGCTCAATCTTC 766  
 QY 560 CGGCGAGAGACACATTAACAGCACTTCAACGAGGCAACACTGCGCATGAGCCGCTGGC 619  
 Db 767 CTGAAGAGTGC-----ACGAGATGGCTGTGTGGGAATTCAGCAACCGAACCCTTACA 820  
 QY 620 TCCTGGAAGGCCCAAGATATGACCATGCGCGCGCTGGCGCTGTACAGCAGGACT 679  
 Db 821 TAGTTGCTCAATAATTTCTCTTGGCACTACTGACAGTTGCAAGATACCGTACGAAT 880  
 QY 680 TTGGCCCTGGCAAAAGGGCCAGATTCGGCATTCCTGCTCAACGGGACTATATGAGCCCT 739  
 Db 881 ATCAGGCTGTCAAGAAAGGTAAAGTTCGGAATAGTCTTGACTTCACTGATACAGGCTC 940  
 QY 740 GGGACAGCAATGAGCTCGGAGCAAGAGGCTGTGAGCAACGGATGGAATTTTCACTTG 799  
 Db 941 T----TACAACTCAGCTGATGACCAACAGCAGCCAAAGAGCGGACTTCCACATTG 997  
 QY 800 GCTGTTTGGCAATCCCATCTTTTGAAGAAGACTATCCAGAGAGCATGAAGAAGCAGC 859  
 Db 998 GCTGTTTGTGATCC--ATTGATAACGAGCACTATCCACAGATATATCAAGATTCG 1054  
 QY 860 TGGGCGAGAGGCTTCCAGCCCTCACTCCCGGACTTGGCATTCCTCAATGCCGAGAGA 919  
 Db 1055 TGAAGGAGAGGCTCCAGGTTCACTCTGAGCAGGCTAAACTGTTAAAGGGCTCGGAG 1114  
 QY 920 CCGACTTCTAGGAGATTAAC---TACACATCCAGATTGCGCGCCACCTAGACGCTC 976  
 Db 1115 ACTACATCGGATTCAGAGATGACATCCAGCTACAGAAAGGCGCAGAACTGTCTCAGC 1174  
 QY 977 CCGTCCCGAGAGCGACTATCTCGGCGCATTCATGAGCAACAGAGAAATGAAGACGCA 1036  
 Db 1175 TGGCGCCAGTAGCTACTGCGGATTCGAGGTTCAATATGTTTTTGCACGCAATGGCA 1234

QY 1037 GCCCGTTGGGAGAGAGCGGCTCGCTGCTGCGCTCTGCGGACATGTTCCGGA 1096  
 Db 1235 AACCGATTGAGACACAGCGCAATTTCTAAGGCTCTACATCGGCCGACGGGATGACG 1294  
 QY 1097 AGCATCTCGCCCGGGGTGTACGCGCTGTACGCAAGC---CCATCTACATCAGAGAACG 1153  
 Db 1295 GGTGCGTAACCTACCTCAAGAGAGATGATGGAATCCAGATCTACATACAGAGAACG 1354  
 QY 1154 GATGCCCGTCCCTTGGAGAGAGAACATGACGTGCGAGAGGCGCTCAAGACCCCTTCC 1213  
 Db 1355 GAATGACCAAGCTTGA-----AACTGACCGGAGCCAGTACTGCGGACGCCACGA 1408  
 QY 1214 GCATCCGTTACTTGTGACTCGCATCTTGAGTCAATTTCAAGGCACTTACCCAGAGCGCG 1273  
 Db 1409 GGGTGGCGTTCTACAGAGAGTACATCGGCGACGTGAAGAGGCCATAGACACG---CGAG 1465  
 QY 1274 TCGTGTCAAGGGGTACTTTGCGTGGCGCTTGTCTGATTAATTGAATGTCAGATGCT 1333  
 Db 1466 CGAAGTGGCTGCTACTTCTGCTGCTCTCGACCAACTTCGAGTGGCTGGCAGGGT 1525  
 QY 1334 ACGGACCCAGATTCCGCGTTCAGCTTACAGACTACACACCTTCAGCCGACCCCAAG 1393  
 Db 1526 ACTGTCGAAGTTCGAGCATGCTTACGTGACTTCAACAGCTCGAACGCCACCCGAAG 1585  
 QY 1394 AGTCTGCCCTGTCTCTCAAGACATGTTT 1422  
 Db 1586 GCTCGCTACTGCTTCAAGGACATGCTT 1614

RESULT 8  
 US-10-425-114-33175  
 : Sequence 33175. Application US/10425114  
 : Publication No. US20040034888A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Liu, Jindong  
 : APPLICANT: Zhou, Yihua  
 : APPLICANT: Kovalic, David K.  
 : APPLICANT: Screen, Steven E.  
 : APPLICANT: Tabaka, Jack E  
 : APPLICANT: Cao, Yongwei  
 : TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 : TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 : FILE REFERENCE: 38-21(5313)B  
 : CURRENT APPLICATION NUMBER: US/10/425, 114  
 : CURRENT FILING DATE: 2003-04-28  
 : NUMBER OF SEQ ID NOS: 73128  
 : SEQ ID NO 33175  
 : LENGTH: 1932  
 : TYPE: DNA  
 : ORGANISM: Zea mays  
 : FEATURE:  
 : OTHER INFORMATION: Clone ID: UC-ZMPLMO17071E02\_FLI  
 : US-10-425-114-33175

Query Match 15.6%; Score 227.4; DB 12; Length 1932;  
 Best Local Similarity 52.2%; Pred. No. 8.4e-57;  
 Matches 735; Conservative 1; Mismatches 637; Indels 36; Gaps 9;

QY 20 TGCCCAACGACTTGAATGGGGCTTCGCAACGGCCGCTACAGATCGAAGGCCCGCTCA 79  
 Db 237 TCCCAAGGGGTTCTGTTTCGGGACGGGACGTCGGGTACAGAGTCGAGGGCGCCGCT 296  
 QY 80 AAGAAGTGGCCCGCGCCGCTTCATCTGGGACAGTACTGCGACCTGGAGCCATCGGCA 139  
 Db 297 CCAACCAACGGCCCGGGCCCTCCATCTGGGATTCATTCGCGCACGTCCTCCAGAAATATTG 356  
 QY 140 CCAACGGCGCAACGGCGATGGCTTGGCATCTACCAACCGCTACAGATGAGGACTTTG 199  
 Db 357 CAGGGAATCAAAATGAGAGACCTTGCAATGATCAATACATCGCTTCAAGAAAGACGTCG 416  
 QY 200 ATCTTTGACCAAGTACGCGGCAAGGCTTCTCTTGTGCTGTGCTGCGGATCA 259

```
Db 417 ATCTCATGAAAAAGTTGAACTTTGATGCTTACCGGTTCTCATCTGATCCAGATCT 476
Qy 260 TTCCCTCGGCGGAGCTTGATCCGTCACAGAGAGGAAATGATTTTACAGCAAC 319
Db 477 TCCCGATGAGGAGG-----AAAGTCATTCAGAAAGGTGTACGATATTAATAATT 530
Qy 320 TGATTTAGCGCTTTGAGCGGGGTATCAGCCTTGGGTACTTTGTACCTCGGATC 379
Db 531 TGATTAACCTATCTCTTCAGCAAGCATGACTCTTCACTCAACCTTTACCTATATATC 590
Qy 380 TGCCCTCAGGCGCTTACGATCGCTATGAGGCTGACTCAACGTGAAAGAGGTCCAGCTG 439
Db 591 TTCTCTTGGCTTGAAGAAATATGAGGAGGTGTTAAGCGCAAGATGCGGACTTG- 649
Qy 440 ACTTTAGCGGTATCGAGGTGGCTTTGAAAGCTTTTGGGACCGAGTCCAGACTGGA 499
Db 650 -TTTACGACTATGCTGACTTCTGTTTAAAGCTACGCGGATGCGCTAAAGACTGCT 707
Qy 500 TCACCATCAAGAMCCCTGGAATTCAGGCACTATGATGATGCAACCGGAGCAACGCC 559
Db 708 TTACATTAATGAGCAAGATAGTAGGCTACTTGGCTATGACACAGGCTCAATCTTC 767
Qy 560 CGGCAAGAGCAGCATTAACAAGCACTCCACGAGGCAACTGCGCACTGAGCGCTGCG 619
Db 768 CTCGAAGGTGC-----ACGAGATGCGCTGCTGGTGGGAATTCAGCAACCGAATCTTACA 821
Qy 620 TCGCTGAAAGGCCAGATCATGAGCCATGCCCGCGCTGGCGCTCTACAGCAGGAGCT 679
Db 822 TAGTTGCTCATTAATTTCTCTTGGGCAATGCTACTGCAAGTTGCAAGATACCGTACGAAT 881
Qy 680 TTCCGCCCTCGCAAAAGGCGCAGATCGGATCTGCTCAACGGGCACTACTATGAGCCCT 739
Db 882 ATCAGGCTGCTCAGAGGTTAAGTTCGAAATATGCTTGAATTCATCTGATGATGAGGCTC 941
Qy 740 GGGACAGCAATGAGCTCGGCAAGAGGCTGTGAGCGAGCGAGTGAATTTCACTTG 799
Db 942 T--TACAACTCACCTGATGATGACAGCAGCGCCCAAGAGCGGAGCTTCCACATTTG 998
Qy 800 GCTGTTGGCCATATCCCATCTTCTTGAAGAAAGATATCCAGAGCAATGAAAGAGCAGC 859
Db 999 GCTGTTGGTGTATCC--ATTGATTAACGAGCACTATTCACAGATTAATGCAAGATCTCG 1055
Qy 860 TGGGAGAGGCTTCCAGGCTTCACTCCGCGAGACTTTGCAATCCCTCAATGCGGAGAGA 919
Db 1056 TGAAGAGAGGCTGCCAGGTTCACTCTGAGCAAGGTTAACTGATGAAGGCTCGGAG 1115
Qy 920 CGACTTCTACGCGATGAATTAC--TACATATCCAGTTGCGCGGCCACTGAGCGGTC 976
Db 1116 ACTACATCGGTATCAACGAGTACACATCCAGCTACATGAAAGGCGAGAGCTGTGTCAGC 1175
Qy 977 CCGTCCCGGAGAGCACTATCTCGGCGCATCCATGAGCACCAGAGAAITAGACGGCA 1036
Db 1176 TGGCGCCAGTACTACTGCGATGGAGAGTTCAATATGTTTGTGACGCAATGGCA 1235
Qy 1037 GCCCTTGGGAGAGAGAGCGGCTCGCTGCTGCGCTCTGCGGACATGTTCCGGA 1096
Db 1236 AACCATTTGGACACAGGGAATTTAAGTGTCTACATCGCCCGAGCGGGAATGTAGC 1295
Qy 1097 AGCATTCGCCCGGGGTGTACGCGCTGTACGCAAGC--CAATTTACATCAGCAGAAAG 1153
Db 1296 GGTGCGTGAATCACTCAAGAGAAATATGGAAATTCAGATCTACATTAACGAGAAAG 1355
Qy 1154 GATGCCCGGCTTGGAGAGAGAAATGACGTGAGAGAGCGGCTCAACGACCCCTTCC 1213
Db 1356 GAATGAGCAGCTGGA-----AACTTGAACCGAGACAGTAACTGCGCGAGCGCAGCA 1409
Qy 1214 GCATCGGTAATTTGAATCGCACTTGACATTTTCAAGGCCATTTACCAAGAGCGCG 1273
Db 1410 GGGTGGGTTCTTACAGAGGTAATCATCGGCAAGTGAAGGCAATGACAG--GAG 1466
Qy 1274 TCGTGTCAAGGGGTACTTTGCGTGGCGGTGCTGCATTAATTTGGAATGTCAGATGCT 1333
Db 1467 CGAAGCTGGCTGCTACTTGTGCTGTCTCTTCGACAACTTGAAGTGGCTGGCAGGGT 1526
```

```
Qy 1334 AGGACCCAGATTTGGGCGTACAGTTCAAGACTACACACCCCTCAAGCGCAGCCCAAGA 1393
Db 1527 ACTGTTCAGATTCGGCATGCTTACTGTGACTTCAACACGCTCGAACCCGACCAAG 1586
Qy 1394 AGTGTCCCTGTGCTCTCAAGACATGTTT 1422
Db 1587 GGTGGGCTTACTGTTCAGGAGCATGCTT 1615
```

```
RESULT 9
US-10-425-114-25892
; Sequence 25892, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovacic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ. ID NOS: 73128
; SEQ. ID NO 25892
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3960-023-C7_FLI
US-10-425-114-25892
```

```
Query Match 15.1%; Score 219.2; DB 12; Length 1737;
Best Local Similarity 51.3%; Pred. No. 2.2e-54;
Matches 727; Conservative 1; Mismatches 654; Indels 36; Gaps 8;
```

```
Qy 20 TGGCCAGCACTTTGAATGGGGCTTCCGAACGCGCGCTTACCAAGTGAAGCGCGCTCA 79
Db 228 TTCCCAAGGGTGTGTGTGCGGAGCGCGCGCTGCGCTACAGAGTGAAGGCAATGCGCG 287
Qy 80 AAGAAGTGGCGCGGCGCGCTTCATCTTGGAGCACTACTGCGACCTGAGACATCGCGCA 139
Db 288 ACAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 347
Qy 140 CCAACGGCGCCAAAGGCGATGTGGCTTGCATCACTCAACCGCTACAGATGAGGACTTTG 199
Db 348 CAACCAAGCGCCACCGCGAGCTGTGTTGAGAGTACATCGCTTACAGAGAGACGTGA 407
Qy 200 ATCTTTGACCAAGTACGCGCGCAAGGCTTACCGCTTCTCTTGTGTGTGCGGATCA 259
Db 408 ATATCATGAAAGATATGAGGGTTGACCGCTTACCGGTTTTCATCTTGTGTCAAGAAATAT 467
Qy 260 TTCCCTCGCGGAGCGCTGATCCCTGCAACGAGAGAGAAATGAGTTTACGCAAAC 319
Db 468 TCCCAATGGAAC-----GAGAAATGAACTGGAAGAGTGGCGCTTACCAACAGGC 521
Qy 320 TGATTTAGCGCTGTGAGCGGGGTATCAGCGCTTGGGTGACTTTGTACCATGAGATC 379
Db 522 TGATTAATTAATGTGTAAGAAAGCATCACCTTACGCAACCTGTACACTACAGACC 581
Qy 380 TGCCCTCAGCGCTTACAGATGCTATGAGGCTGCGCTCAACGTGGAAGAGGTCCAGCTGG 439
Db 582 TACCGAGAGCGCTGAGGTTCGGTACGAGAGGCTGTGTGAGCAGAGAAATGTGT---GAGAT 638
Qy 440 ACTTTGACCGGTATGCAAGGTGTGCTTTGAAAGCTTTTGGGGAACGAGTCCAGACTGGA 499
Db 639 CGTTGCGAGACTTACGCGCACTTGTGCTTGGGGGCTTGGGAGACAGGTTGAAGAACTGGC 698
Qy 500 TCACCATCAAGAMCCCTGATTCAGGCACTTATGATGATGCAACCGGAGCAACGCGCC 559
```

```

Db      699 TGAAGTTCACGAGCCCGGGGTGTGGCCGCTTGAAGGTACGACGACGAGGTTCCGCG 758
Qy      560 CGGACGAGGACGAGATTACAGAGACTCCACGAGGGCAACACTGCCACTGAGCCGCGG 619
Db      759 CGGAGAGGTTCACGAGGTGCGAG-----GCCGGGGGGGACTGGGACACGAGGCTTAAG 812
Qy      620 TCGCTGAAAAGGCCAGATCATGAGCCATCCCGCGCGCTGGCGCTTACAGACGAGGACT 679
Db      813 TCGTGGGACGACACTCATCTCTCCACGCCCGCGCTGTCAGAGGTACCGCGCAGGC 872
Qy      680 TTGCCCCCTTGGCAAAAGGGCCAGATCGGACTTTCGCTCAACGGGACTACTATGAGCCCT 739
Db      873 ACCAGCCGACGCGAGAGGGGCGAGGTCTCGGATCTCTGCTGAGATTTCGTGTGTACGAGCCC 932
Qy      740 GGGACAGCAATGAGCTCTGGGACAGAGGCTGTGAGCGACGAGTGAATTTCACTATG 799
Db      933 TACGGCGGAGCTAGCGCGGACCGGCGCCCGCTCAAAAGTCCAGAGACTTCCACGCTCG 992
Qy      800 GCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAGAAGCAGC 859
Db      993 GATGTTCTCGACCCCATCTGTCT---ACGGCGAGTACCCCAAGTGGTCCGGAGAAAGC 1049
Qy      860 TGGGCGAAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGCCATCTTCATATGCCGAGAGA 919
Db      1050 TCAAGGGGACGGCTCCCAAGTTTCAAGGCTGAGGA---GGCCGGTCTAGTCCGAGGCTCCA 1106
Qy      920 CGGACTTTTACGAGCATGAATTTACTACATCCAGTTCCGCGCGCCACTAGACGAGTCCCG 979
Db      1107 TCACACTAGCTGAGATCACAGATACCTGCTACTACGTCGAGTACGAGCGCAAAACG 1166
Qy      980 TCCCGAGAGCG-----GACTATCTCGGCGCCCATCATGAGACACAGAGAAATTAAG 1030
Db      1167 CTACGGGGGGGCGCGCCAGCTACTCGTCCGACTGGACGCGTGAATTCGTATAGAACGCG 1226
Qy      1031 ACCGACAGCCCGTTGGCGAGAGAGCGGCTCCGCTGGCTGCGCTCTGCGCGACATGT 1090
Db      1227 ACCGTTGCGGATTTGGACCAAGGGCGCAACTCAGACTGCTCTACATGCTCTTGGGGAC 1286
Qy      1091 TCCGGAAGACATTCGCGCGGGGTGATCGGCGCTGTACGGCAGAGCCCATTAATCACGAGA 1150
Db      1287 TGTACAAAGCCGTACCTTACGTACAGAGAGATACGCAACCCCA---CGATGCTCTGT 1343
Qy      1151 ACAGATGCCCGTGCCTCGAGAGAGAACTGACGTGCGAGAGCGCGCTCAACGACCCCT 1210
Db      1344 CAGAGAACGGTATGAGACGACCCGGGCAACCTCACGTTGGGCGAGGGGTCCGGACCGCG 1403
Qy      1211 TCCGCAATCCGTAATCTTGAATCTGCACTTGAGCTGATTTCCAAAGCCATTACCGAGAGC 1270
Db      1404 CGAGGGTGGCTTACT---ACCGAGGCTACGTCGCCGAGCTGAAGGCGGCGATCGACGCG 1460
Qy      1271 GCGTCTGCTGCAAGGGGTACTTTTGGCGGGCTTGGCTGCTGATTAATTGGAATGTCAGATG 1330
Db      1461 GCGCAACTGCGGAGATTAATCTGCTGCTGCTGCTGCTGCACTTGAATGGAAGCTGG 1520
Qy      1331 GCTACAGGACCCAGATTGCGGCTGACGTTACAGACTACACACCTCAACGCGACGCGCA 1390
Db      1521 GGTACACGTCCTCGGTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1580
Qy      1391 AGAAGTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1428
Db      1581 AGAGCTCGGCGTACTGTTCAAGGATGTCAATCGCCGCG 1618

```

RESULT 10  
US-10-425-114-13710

; Sequence 13710; Application US/10425114  
; Publication No. US2004003488A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

```

; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13710
; LENGTH: 1759
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-030-G7_FLI
US-10-425-114-13710

```

Query Match 12.9%; Score 187.6; DB 12; Length 1759;  
Best Local Similarity 51.0%; Pred. No. 5e-45;  
Matches 663; Conservative 1; Mismatches 600; Indels 36; Gaps 8;

```

Qy      138 CACCAACGGCGCCACGAGATGAGCTTGGCTTGCATCACTACACCGCTACGATGAGACTT 197
Db      368 CGAAACAAACGCCACCGCGAGCTGACTGTTGACGATACATCGCTTACAGAGAGACT 427
Qy      198 TGATCTTTGACCAAGTACGCGCAAAAGGCTTACCGCTTCTCTGTGTGTGGAT 257
Db      428 GAATATCATGAAGATATGAGGTTTCAGACGCTTACCGGTTTTCATCTTGTGCAAGAT 487
Qy      258 CATTCCCTTCGCGGAGGCTGATCCCTCAACGAGAGAGGAATGAGTTTACAGCA 317
Db      488 ATTCCCAATTTGAACCGAGAA-----GTGAATCGAAAGAGTGGCGTACTACACAG 541
Qy      318 ACTGATTAAGCGCTGTTGAGGCGGGGTATACGCGCTTGGGTGACTTTTACCACTGGA 377
Db      542 GCTGATTAATCATGATGTGAAGAAAGCATCACCTTACGCAAACTGTACCACTACGA 601
Qy      378 TCTGCTTACAGCGCTTACAGATGCTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCT 437
Db      602 CTACCGGAGGCGCTGAGAGGTCCGATACGAGGCGTGTGAGCAGAGAAAGTGT---GAG 658
Qy      438 GGAATTTGAGCGGTATGCGAGGTTGCTTGAACGTTTGGGAGCCGATCCAGAACTG 497
Db      659 ATGTTTCGACAGATCGCGACTTCTGCTTGGGGGCGTTGCGGACAGGGTGAAGACTG 718
Qy      498 GATCACATCAAGAACCCGATGATTCAGGCGCATCTATGATATATGCAACCGCGAGCAACG 557
Db      719 GCTGAGTTTACAGAGCGCGGGGTGTGCGCGCTTACGATGAGACAGCGAGGTTCCG 778
Qy      558 CCGGCGAGGAGCAGATTAACAGACATCCACCGAGGCGCAACTGCACTGAGCGGTG 617
Db      779 GCGGGGAGGTGACGCGGGGTGCGAGG-----CGGGGGGGAATCGGGGACCGAGCCCTA 832
Qy      618 GCTGCTGGAAGAGCCCAATCATGAGCCATGCGCGCGCGCTGAGCCGTATACAGAGGGA 677
Db      833 CGTGTGTGCGCACACACTTCTCTTCCACGCGCGCGCTGTCAGAGGATACCGCGCAG 892
Qy      678 CTTTGGCCCTGCAAAAGGGCCAGATTCGSCATCTGCTCAACGGGCACTACTATGAGCC 737
Db      893 GCAACAGCGACCAAGAGGGGAGGGTGGGATCTGCTGATTTGTGTGTGTAAGAGCC 952
Qy      738 CTGGAACAGCAATGAGCTTCGAGCAAGAGAGGTGCTGAGCGACGAGATGAATTTTCAT 797
Db      953 CTTCAAGGGGAGTACAGCGCGCGACCGGGCGCGCTCAAAAGTCCAGAGACTTTCACGT 1012
Qy      798 TGGCTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGACATGAAGAAGCA 857
Db      1013 CGGATGTTCTTGCACCCCATCTCT---ACGGCGAGTACCCCAAGTCCGCTCCGAGAAAG 1069
Qy      858 GCTGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTTGGCATCTCTCAATGCCGAGAG 917
Db      1070 CTTCAAGGGGAGGCTTCCCAAGTTTACGCGCTGAGGA---GCGCGTCTTACTCCGAGGCTC 1126
Qy      918 GACGACTTCTACGAGATGAATTAATCAATCCAGTTTGCAGCGGCACTTACAGCGGTCC 977

```

Db 1127 CATGACCTACGTGAGAGTCAACAGTACACTGCTTACTACGTGCTGATGACGCGCCAA 1186  
QY 978 CGTCCCGGAGACG-----GACTATCTGGGCGCCATCCATGAGCAGCAGAGAAATA 1028  
Db 1187 CGCTACGCGCGCGCGCGCGCGAGCTACTGTCGACTGGCAGCTGAGTTCGTATGAAAG 1246  
QY 1029 GAGCGGAGCGCGCTTGGCGAGAGAGCGGCTCGCTGGCTGGCTCTGCGCCGAGCAT 1088  
Db 1247 CGACGGGTGCTCCGATGACCAAGGGCGAACTCAGACTGGCTTACATCGTGGCTTGGGG 1306  
QY 1089 GTTCCGGAACGATTCGCGCGCGGTGTAGGGCTGTACGGCAAGCCATTCATCAACGA 1148  
Db 1307 ACTGTACAAAGCCGTCACTTACGTCAAGAGAGATGAGGCAACCCACAGATGCTCTGTC 1366  
QY 1149 GAACGATGCGCCCTGCGCTGAGAGAGAAATGACGTGCGAGAGGCGCGTCAACGACC 1208  
Db 1367 AGAAGACGATATGAC---GACCGGGGCAAGTCACGGTGGGCGAGGGCGCGCAGCGC 1423  
QY 1209 CTTCGCGATCCGCTACTTTGACTGTGCACTTGACTTCGACTTTCGAAGCGCATTAACCGAGA 1268  
Db 1424 GCGCAGGGGTGCTTACT---ACCGAGCTACGTGCGCGAGCTGAAGGGCGGATGACGG 1480  
QY 1269 CGGCGTGTGCTCAAGGGGCTACTTTGCGTGGGCGTTCGATTAATTGATGATGAGA 1328  
Db 1481 CGGCGCCTACCTGCGTGGATATCTTCTGCTGCTGCTGCTGCACTTCGAGTGGAAAGCT 1540  
QY 1329 TGGCTACGAGCCCAAGATTGCGCGTCAAGTTCACAGACTACACACCTCAAGCGCAGCC 1388  
Db 1541 GGGGTACACGTCCGCGTTCGCTGCTGCTACGTGCACTTCAGAGACGCTCCGCGCTACCC 1600  
QY 1389 CAAGAGCTGCGCTCGCTGCTCAAGGACATGTTGGCGGCC 1428  
Db 1601 CAAGAGCTGCGCTGCTGCTCAAGGATGATCAACCGCGC 1640

## RESULT 11

US-10-425-114-24850  
; Sequence 24850, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 24850  
; LENGTH: 1735  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3689-207-B2\_FLI  
US-10-425-114-24850

Query Match 12.7%; Score 184.4; DB 12; Length 1735;  
Best Local Similarity 50.3%; Pred. No. 4.4e-44;  
Matches 710; Conservative 0; Mismatches 666; Indels 36; Gaps 9;

QY 16 GCTCTGCCAAGCATTTTGAATGGGCTTCGCAACGGCGCTTACCAATCGAAGCGCGC 75  
Db 166 GGGTTCGCCGAGCGCTTCTGCTTCGGACCGCGCGCTGACCAAGTCAGAGGAGATG 225  
QY 76 GTCAAGAAAGTGGCGCGCGCGCGCTGCAATCTGGGACACGTAATGCACTGAGACCATG 135  
Db 226 GCCAAGCAAGCG 285  
QY 136 CGCAGCAAGCG 195

Db 286 ATCCCTAACAGCCACCGCTGACGTCGACGTCGACGATATCATCGGTACAAAGAAAT 345  
QY 196 TTTGATCTTTTACCAAGTACCGCGCGCAAGGCTTACCGCTTCTCTGTCGTCGCGCG 255  
Db 346 GTCAACATATGAGAAACATGGGCTTTGATGGTACCGGCTTTTCATCTCTTGTCGAGG 405  
QY 256 ATCATTTCCCTCG 315  
Db 406 ATTTTCCAGATGAACTGGCAAG-----GTGACACGAAAGAGATGATTAACACAAAC 459  
QY 316 AAACGATTTGACGCGCTGTTGAGCGCGGATATACGCTTGGGTGACTTTGATCCACTGG 375  
Db 460 AGGCTCATGATATTAATGATCTCCAGCAAGGATATCGCGCGCTTATGCAATCTTACCACTTAT 519  
QY 376 GATTCGCTTACGCGGCTTACAGATTCCTATGAGAGGTGCTCAACGTGAGAGGCTCAG 435  
Db 520 GACCTCCATTTGGCCTCATGATGAACGTAACCTGGGGCTGGCTTACG---CCMAAGATTGG 576  
QY 436 CTGCACTTTGAGCGGTATGCGAGGTTGCTTTGAACTTTTGGGAGCCGAGTCCAGAAC 495  
Db 577 GAGGCTTTGACACTACCGCGAGTTCTGCTTCCAGCGCTTCGAGACAGGGTGAAGAAC 636  
QY 496 TGGATCAACATCAACGAMCCCTGATTCAGGCCATATGATATGATTCACCGCGCACAC 555  
Db 637 TGGTTTACCTTCAACGAGCGAGGTGCTGCTGCTTGGGCTTACGACATGAGCTTTCAC 696  
QY 556 GCCCGCGCGAGGAGCAGATTTACAGACACTCCACCGAGGCAACCTGCACTGAGCGG 615  
Db 697 GCAACCGGAAAGTG-----TTCGGGTGCTCCGCGGAGGCACTCAACACGAGGCGG 750  
QY 616 TGGCTGCTGGAAGGCGCCAGATCATGAGCCATGCGCGCGCGCGCGCTTACAGCAGG 675  
Db 751 TACCTGTGCGACACCATCTCATCTTCTCATGAGCTGCGGTGACGCGATCCCGGAC 810  
QY 676 GACTTTGCGCCCTTCGCAAAAGGCGCATGCGCATCTCGTCAACGCGCATCTATGAG 735  
Db 811 AAGTATGAGTTTCAACGAAAGGAGATGAAATCTCTGGAATTTCTGTCGTGATGAA 870  
QY 736 CCCGCGGACGACATGAGCTTCGAGCAAGAGGCTGCTGAGCGACGAGATGAAATTTTAC 795  
Db 871 CTTT---CAGCGACAGCATTGGGACAGGCTGAGCAGACAGGAGCAAGGACTTCAC 927  
QY 796 ATTGCTGTGTTTGGCAATCCCATCTTCTTGAAGAAGCATATCAAGAGACATGAAGAAG 855  
Db 928 CTAGCTGTGTTCTTGAACCCCA---TTGTACATGAGCGTACCCGTATCGATGCAAGAG 984  
QY 856 CAGCTGGCGGAGAGGCTTCAAGCGCTTCACTCCGCGGAGCTTGGCATCTCAATGCGGGA 915  
Db 985 ATTGCCAAGACAGGCTTACGCTGTTGTTCAAGATGAAGACAGAGATGCTGAAGGCTCT 1044  
QY 916 GAGACCGACTTCTACGCGCATGAAATTAACAATCCAGTTGCGCGCCACTAGACGCT 975  
Db 1045 ATGACATATGTTGGCATCAACCACTACACTTCTTTACATGAAGAGCCCTGGAGCATGG 1104  
QY 976 CCGTTCGCGAGCGAGATATCTGGCGCATTCATGAGACACGAGAAATAG---GAC 1032  
Db 1105 AACCTGACGCGCATGATGATCAAGATGATTTGATGTTTGTCTGACAGCAAT 1164  
QY 1033 GCGAGCGCGCTTGGCGAGGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1092  
Db 1165 GGAATTTCTATTTGGGCTCAGCAAGAACTTCACTGCTGATGATTTGGTGGGAGCATC 1224  
QY 1093 CGGAGCATCTGCGCGCGGTGTACGCGCTGTAGCGCAAGC---CATTTACATACCGAG 1149  
Db 1225 AAAGAGCTGTCAAGTATGTCAAGAAATTTAAGAAATCTTCAATGATCTTGTGAA 1284  
QY 1150 AAGGATCCCGTGGCTTGGAGAGAGAAACATGACGTGAGAGGCGCTCAACGACGCC 1209  
Db 1285 AACGGAATGACCAACCTGTGATGTCTGATTA---TCAAGGTGTGATGACACA 1338  
QY 1210 TTCCGATCCGTAATTTGATCTGCACTTGAATCTGATTTTCAAGGCCATTAACCGAGAC 1269

Db 1339 GTAAGATCCGTTA---TTACAGAGACTACATTAAGTCAAGGAAAGCAATAGATAT 1395  
Qy 1270 GGGGTGTCGTCAAGGGGTACTTTGGTGGCGTGTGCTGATACTTGAATGTCAGAT 1329  
Db 1396 GGTGCCAGAGTCTTGGGTACTTTGGTGGCGTGTGCTGATACTTGAATGTCAGAT 1455  
Qy 1330 GGGTCCAGAGTCTTGGGTACTTTGGTGGCGTGTGCTGATACTTGAATGTCAGAT 1389  
Db 1456 GGGTCCAGAGTCTTGGGTACTTTGGTGGCGTGTGCTGATACTTGAATGTCAGAT 1515  
Qy 1390 AAGAAGTCCGTCCTGCTCTCAAGAGCATGTT 1421  
Db 1516 AAGGACTCAGCTTCTGCTTCAAGCATATGCT 1547

RESULT 12  
US-10-425-114-27741  
; Sequence 27741, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 27741  
; LENGTH: 1757  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB4739-015-E1\_FLI  
US-10-425-114-27741

Query Match 12.7%; Score 184.4; DB 12; Length 1757;  
Best Local Similarity 50.3%; Pred. No. 4,4e-44;  
Matches 710; Conservative 0; Mismatches 666; Indels 36; Gaps 9;  
Qy 16 GCTCTGCCCAAGCACTTTGAATGGGGCTTGCACACGGCGCCCTTACCAAGATGGAAGGCC 75  
Db 188 GCGTTCCTCCGAGCGCTTCTTCTGAGGACGCGCGCTGCGTCAAGGTCGAGGGATG 247  
Qy 76 GTCAAGAAAGTGGCGCGCGCGCTTCCATCTGGGACACGTAAGTCCACCTGAGCCATCG 135  
Db 248 GCCAAGCAGCGCGCGCGCGCGCGCTTCCATCTGGGACACGTAAGTCCACCTGAGCCATCG 307  
Qy 136 CGCAGCAACGGCGCCCAAGCGCGATGTCGCTTCCGATCACTACCAACGCGCTACGATGAGAC 195  
Db 308 ATCCCTTAACATGCAACCGCTGACGTGACGATGACGATGATCATCGTCAAGAAAGAT 367  
Qy 196 TTGTGATCTCTTGAACAAAGTACGGCGCAAGCGCTTCCCTTCTTGTGCTGCTGCGG 255  
Db 368 GTCAACATAATGAAGAACATGGCTTGTATGCGTACCGGTTTCTGATCTCTTGTGTCAGG 427  
Qy 256 ATCATTCCTCCCTCGCGCGAGCTGATCCGTCACAGAGAGGAATTGATTTACAGC 315  
Db 428 ATTTTCCCAATGGAATGCGCAG-----GTGAACAGAGAGAGAGTACTACTACAC 481  
Qy 316 AAACTGATGACGCGCTGTTGAGCGGGGTATCAACGCTTGGTGAATTTGATCACTGCG 375  
Db 482 AGGCTCATAGATTAATGATGCTCCAGCAAGTATCGCGCTGATGCAAACTCTACATTAAT 541  
Qy 376 GATTCGCTCAGGCGCTTCAAGATGCTGATGAGAGGCTGCTCAACGTAAGAGGTCAG 435  
Db 542 GACCTCCCATTTGGATGCTCAAGAAAGTACCTGGCTGCTGAGC---CCAAAGATTGTG 598  
Qy 436 CTGAGCTTTGAGCGGTATGCGAGGTTGTGCTTGAACGTTTGGGAGCCAGTCCAGAAC 495

Db 599 GAGCGCTTTGACAGACTACCGCCGAGTTCTGCTTCCAGCGCTTGGAGACAGGGTGAAGAAC 658  
Qy 496 TGGATACCATCAACAGAACCTTGGATTTAGAGCCATCTATGATATATGCAACCGGACAGAAC 555  
Db 659 TGGTATACCTTCAACAGAGCGAGGTCGTGCTTCTGAGCTTACACAAATGAGCTTGCAC 718  
Qy 556 GCCCGGAGAGAGAGCAATTAACAAGCACTCCACCGAGGCAACAATGCACTGAGCCG 615  
Db 719 GCACCGGAAAGTG-----TTCCGGGTGCCCCCGGAGGCAACTCAACAGAGCCG 772  
Qy 616 TGGCTGCTGAGAAAGGCCAGATCAAGCATGAGCATGCCCCGCGCTGAGCTTACAGCAGG 675  
Db 773 TACCTTGACACACCATCTCATCTTCTCATGACAGCTGCGGTCAAGGAGTACCCGAC 832  
Qy 676 GACTTTCGCCCTCCGAAAGGCCAGATCGATCGATCTGCTCAACCGGCACTACTATGAG 725  
Db 833 AAGTATCAGCTTCAACAGAAAGGGAAGATTGAAATTCCTGGAATTCGTGTGTCGAA 892  
Qy 736 CCTGGGACAGCAATGAGCTCGGACAAAGAGGCTGTGAGCGAGATGAAATTTAC 795  
Db 893 CTTT---CAGCGACAGCAATGCGGACAGGCTGACAGCACAGGACCGCAAGCTTCCAC 949  
Qy 796 ATTGCTGTGTTGCAATCCCATCTTCTGAAGAGCACTATCAAGAGCATGAAGAG 855  
Db 950 CTAGGCTGTTCTTGAACCCA---TTGATCATGACGAGTACCGTACTGATGCAAGAG 1006  
Qy 856 CAGCTGGGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTCCATCTCCATGCGGAG 915  
Db 1007 ATTGCCAAGAACAGGCTACCGTGTGTTGATGAGATGAAGAACAGATGGAAGGCTCT 1066  
Qy 916 GAGACGACTTCTACGCGATGAATTAATCACTACATCCAGTTCGCGGCCACTAGACGT 975  
Db 1067 ATGAGCTATGTTGGATCAACCACTACACTCTTCTTACATGAAGAACCTGGAGACATGG 1126  
Qy 976 CCGCTCCCGAGACGAGCTATCTGCGCGCCATCAATGAGACCAAGAAATGAG---GAC 1032  
Db 1127 AACCTGACGCACTACAGTACAGATGATGATGATGATGATGATGATGATGATGATGAT 1186  
Qy 1033 GCGAGCCCGCTTGGCGAGAGAGCGCGCTGCGCTGCGCTGCGCTGCGCGAGATGTTG 1092  
Db 1187 GAGATTCATATGCGCTGCTACCGCAAACTCTTCTGCTGATGATGATGATGATGATGATG 1246  
Qy 1093 CGAAGCATCTCGCGCGGCTGATGCGGCTGATGCGGAGC---CCTATCTATCAACGAG 1149  
Db 1247 AACAGAGCTCTCACTATGTCAGAGAACTTACAAAAATCTTCAATGATCTTCTGAA 1306  
Qy 1150 AACGATGCGCGCTGCGAGAGAGAACTATGATGCGAGAGGCGCGTCAACGACCCC 1209  
Db 1307 AACGATGAGACCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1360  
Qy 1210 TTCCGATCGGATCTTGAATCTGCACTTGCATGATGATGATGATGATGATGATGATGAT 1269  
Db 1361 GTAAGATCCGTTA---TTACAGAGACTACATTAAGTCAAGAGCAATGATGAT 1417  
Qy 1270 GCGCTGCTGCAAGGGGTACTTTGGTGGCGTGTGCTGATTAATGATGATGATGATGAT 1329  
Db 1418 GGTGCCAGAGTCAATGGGTACTTTGGTGGCGTGTGCTGATTAATGATGATGATGATGAT 1477  
Qy 1330 GCGTCCAGAGCCAGATTCGCGCTCAAGTTCACAGACTACCAACCTCAAGCGCACGCC 1389  
Db 1478 GGGTACACTTCCGCGTTTGGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 1537  
Qy 1390 AAGAAGTCCGTCCTGCTCTCAAGAGCATGTT 1421  
Db 1538 AAGGACTCAGCTTCTGCTTCAAGCATATGCT 1569

RESULT 13  
US-10-369-493-31938  
; Sequence 31938, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:



Query Match 12.6%; Score 182.8; DB 9; Length 1734;  
 Best Local Similarity 49.2%; Pred. No. 1,3e-43;  
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

18 TCTGCCCAAGCATTTGATGAGGCGCTTGGCAAGCGCCCTACCAATGAGAGCGCGT 77  
 Db TTTTCTGATGATTTTCACTTTGGGACAGCTGCTCGCGCTTTCAGTACGAAGGTGCAAC 149  
 QY 78 CAAGAAGGTGGCGCGCGCGCTTCCATCTGGACAGTACTGCGCACTGAGCCATGCGC 137  
 Db 150 AAGTGAAGGTGGCAAGTCTCAACTATATGAGATCACTTCAAGCTTCAAGTACAGAAAG 209  
 QY 138 CACCAACGCGCGCAAGCGCGATGCTGATCACTTACACCGCGATGAGAGCTT 197  
 Db 210 GACCAAAATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 269  
 QY 198 TGATCTTTGACCAAGTACGCGCGCAAGCGCTTCCCTTCTGCTGCTGCTGCTGCT 257  
 Db 270 AAATTTGATGAGAGCTTAAACATGACGCTTCCGATTTTCAATCTCGTGTCAAGATT 329  
 QY 258 CATTCCCTTGGCGGCGAGGCTGATCCCTTCAACGAGAGAGGAATTGATTTTACAGCA 317  
 Db 330 AATACCCAGTGAAGAGCTTAAAGATGAGATGAACAAAGAAAGTGTACATTTCAAGAGA 389  
 QY 318 ACTGATTTGACGCGCTGTTGAGCGCGGATACGCGCTTGGGATCTTGTACCACTGGA 377  
 Db 390 TCTCATATGACGAATCTTCTTCTATATGATATACATCACTTTCATGACGCTTATCTATG 449  
 QY 378 TCTGCTTCAAGCGCTTCAAGATGCTATGAGAGGCTGCTCAAGTGAAGAGTCCAGCT 437  
 Db 450 CCACCCACATCTTTGAGAGAGCAATATGCTGCTTCTTAAAGCTTAAATTCGATGA--- 506  
 QY 438 GGACTTTGAGCGGATGCGAGGTTGCTTGTAAAGCTTTGGGAGCCGATCCAGAACTG 497  
 Db 507 AGATTTTCAGATTTTGGCAAGAAATTTGTTGAAAGATTTTGGAGATTAAGTTAAGATG 566  
 QY 498 GATCACCATCAAGCAAGCCCTGATTTGAGCGCATCTATGATATGATGATGATGATGAT 557  
 Db 567 GACCAACATTAAGCAAGACCTTATATATGATGATGATGATGATGATGATGATGATGAT 626  
 QY 558 CCGCGGAGAGAGC-----AGCATTTAACAAGCACTCCACGAGGCAACTGCGCACTGA 611  
 Db 627 GCGTGAAGATGCTCAAAATGGGTAACGAAGATGTCAGAGGATGATGATGATGATGAT 686  
 QY 612 GCGGTGAGCTGCTGGAAGAGCCCAAGATCATGAGCATGCGCGCGCTGCGCTTACAG 671  
 Db 687 GCGTTACATTTGTTTCAATCACAACCTTCTTGGCCCATGCGCTGCAAGTGAAGAAATTCG 746  
 QY 672 CAGGAGCTTTGCGCCCTCGCAAAAGGCGCATGCGCATCTCGCTCAAGCGGCACTACTA 731  
 Db 747 AAATGTGAAAAAACTTGGCATGATGCGCAAAATGGGATGATGATGATGATGATGATGAT 806  
 QY 732 TGAGCCCTGAGAGCAAGATGAGCTCGGACCAAGAGAGCTGCTGAGCGAGTGAAT 791  
 Db 807 CGAGCCTTATCATTTCCGATTTCACTGACGATTAAGAGAGAGCTGAAAGAGACTCTTGCCT 866  
 QY 792 TCACTTTGCTGTTGCTGCTCAATCTTCTTGAAGAGAGCTATCAAGAGAGATGAA 851  
 Db 867 TGAATTTGAGTGCATCTTATCAGATCATTC--ACGAGAGATTTATCAAGAGATTTGAAA 923  
 QY 852 GAAGCAGCTGGGAGAGAGGCTTCCAGCCCTCACTCCGCGGAGCTTTGCCATCCCAATGC 911  
 Db 924 AAGGTAAGCGGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 983  
 QY 912 CGAGAGAGCGAGCTTCTACGAGCATGATTAATTAATTAATTAATTAATTAATTAATTA 971  
 Db 984 TTCA---TCAGATTTGCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1040  
 QY 972 CGGTCCCTGCGCGAGAGCGAGCTATCTGCGCGCATCATGA-----GCAACAGGA 1022  
 Db 1041 TCACATTAAGACCTTGAAGAAACCTGTTCAAAATGACACCATGATGTGAATGAAACTGAC 1100

QY 1023 GAATTAAGACGAGCGCCCTTGGCGAGAGAGAGCGGCTTCCGCTGCGCTTCCGCC 1082  
 Db 1101 TAATCAAGTGTGATCATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1160  
 QY 1083 GGAATGTTCCGAGAGCATTCGCCCGGCTGATGCGCTGTAC--GCAAGCCATCTA 1139  
 Db 1161 GGAAGCTTACGAAAGTCTTAACTATATTAAGAGAGATTAATTAATTAATTAATTAAT 1220  
 QY 1140 CATCACGGAAGAGATGAGCGCGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1199  
 Db 1221 CATCAAGAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1280  
 QY 1200 CAACGAGCCCTTCCGATCGATCTTGTGATCTGCACTTGTGATCTGATTTTCAAGGCA 1259  
 Db 1281 GAAGACACATTTAGATTTGAATACATTAAGACACATTTGAGAGATTTTCAAGAGTAT 1340  
 QY 1260 TACCAAGAGCGGCTGCTGATCAAGGAGTATCTTGGTGGCGCTGCTGATTAATTTGA 1319  
 Db 1341 AGTGAAGATGAGGTGTGAGCTTAAGAGATTAATTAAGATGCTGATGATGATGATGAT 1400  
 QY 1320 ATGTCAAGTGTGATGAGAGCCGATTTGG 1349  
 Db 1401 ATGGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1430

## RESULT 15

US-09-938-842A-1282  
 ; Sequence 1282, Application US/09938842A  
 ; Publication No. US2004009476A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Krepe, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE OF INVENTION: SAME, AND METHODS OF USE  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 1282  
 ; LENGTH: 1734  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 ; US-09-938-842A-1282

Query Match 12.6%; Score 182.8; DB 11; Length 1734;  
 Best Local Similarity 49.2%; Pred. No. 1,3e-43;  
 Matches 664; Conservative 1; Mismatches 658; Indels 27; Gaps 6;

18 TCTGCCCAAGCATTTGATGAGGCGCTTGGCAAGCGCCCTACCAATGAGAGCGCGT 77  
 Db 90 TTTTCTGATGATTTTCACTTTGGGACAGCTGCTGCGGCTTTCAGTACGAAGGTGCAAC 149  
 QY 78 CAAGAAGGTGGCGCGCGCGCTTCCATCTGGACAGTACTGCGCACTGAGCCATGCGC 137  
 Db 150 AAGTGAAGGTGGCAAGTCTCAACTATATGAGATCACTTCAAGCTTCAAGTACGAAG 209  
 QY 138 CACCAACGCGCGCAAGCGGATGCTTGGCATCACTACCAACGCTTACAGATGAGACTT 197  
 Db 210 GACCAAAATGATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 269  
 QY 198 TGATCTTTGACCAAGTACGCGCGCAAGAGCTTCCGCTTCTCTTGTGCTGCTGCGGAT 257  
 Db 270 AAATTTGATGAAGAGCTTAAACATGAGAGCTTTCGATTTTCAATCTCGTGTCAAGATT 329  
 QY 258 CATTCCCTGCGCGGAGGCTGATCCCGTCAACGAGAGAGGAATTGATTTTACAGCAA 317





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 23:31:59 ; Search time 5013.54 Seconds  
(Without alignments)  
10211.501 Million cell updates/sec

Title: US-10-026-140-3  
Perfect score: 1455  
Sequence: 1 atgccgcagctcctagctctct.....99gcttaagctgcgcataa 1455

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Pending\_Patents\_NA\_Main:\*

- 1: /cgn2\_6/ptodata/2/pna/US06\_PCTUS\_COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/pna/US08\_COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq:\*
- 8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pna/US092A\_COMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pna/US092B\_COMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pna/US093A\_COMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq:\*
- 20: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq:\*
- 21: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq:\*
- 22: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq:\*
- 23: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq:\*
- 24: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq:\*
- 25: /cgn2\_6/ptodata/2/pna/US096A\_COMB.seq:\*
- 26: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq:\*
- 27: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq:\*
- 28: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq:\*
- 29: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq:\*
- 30: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq:\*
- 31: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq:\*
- 32: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq:\*
- 33: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq:\*
- 34: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq:\*
- 35: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq:\*
- 36: /cgn2\_6/ptodata/2/pna/US098D\_COMB.seq:\*
- 37: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq:\*
- 38: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq:\*
- 39: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq:\*
- 40: /cgn2\_6/ptodata/2/pna/US099D\_COMB.seq:\*
- 41: /cgn2\_6/ptodata/2/pna/US099E\_COMB.seq:\*
- 42: /cgn2\_6/ptodata/2/pna/US099F\_COMB.seq:\*
- 43: /cgn2\_6/ptodata/2/pna/US100A\_COMB.seq:\*

- 44: /cgn2\_6/ptodata/2/pna/US100B\_COMB.seq:\*
- 45: /cgn2\_6/ptodata/2/pna/US101A\_COMB.seq:\*
- 46: /cgn2\_6/ptodata/2/pna/US101B\_COMB.seq:\*
- 47: /cgn2\_6/ptodata/2/pna/US102A\_COMB.seq:\*
- 48: /cgn2\_6/ptodata/2/pna/US102B\_COMB.seq:\*
- 49: /cgn2\_6/ptodata/2/pna/US103A\_COMB.seq:\*
- 50: /cgn2\_6/ptodata/2/pna/US103B\_COMB.seq:\*
- 51: /cgn2\_6/ptodata/2/pna/US104A\_COMB.seq:\*
- 52: /cgn2\_6/ptodata/2/pna/US104B\_COMB.seq:\*
- 53: /cgn2\_6/ptodata/2/pna/US106\_COMB.seq:\*
- 54: /cgn2\_6/ptodata/2/pna/US107A\_COMB.seq:\*
- 55: /cgn2\_6/ptodata/2/pna/US107B\_COMB.seq:\*
- 56: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq:\*
- 57: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq:\*
- 58: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq:\*
- 59: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq:\*
- 60: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq:\*
- 61: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq:\*
- 62: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq:\*
- 63: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq:\*
- 64: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq:\*
- 65: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq:\*
- 66: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq:\*
- 67: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq:\*
- 68: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq:\*
- 69: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq:\*
- 70: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq:\*
- 71: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq:\*
- 72: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq:\*
- 73: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq:\*
- 74: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq:\*
- 75: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq:\*
- 76: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq:\*
- 77: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq:\*
- 78: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq:\*
- 79: /cgn2\_6/ptodata/2/pna/US6023A\_COMB.seq:\*
- 80: /cgn2\_6/ptodata/2/pna/US6023B\_COMB.seq:\*
- 81: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq:\*
- 82: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq:\*
- 83: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq:\*
- 84: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq:\*
- 85: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq:\*
- 86: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq:\*
- 87: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq:\*
- 88: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq:\*
- 89: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq:\*
- 90: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq:\*
- 91: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq:\*
- 92: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq:\*
- 93: /cgn2\_6/ptodata/2/pna/US6037\_COMB.seq:\*
- 94: /cgn2\_6/ptodata/2/pna/US6038\_COMB.seq:\*
- 95: /cgn2\_6/ptodata/2/pna/US6039\_COMB.seq:\*
- 96: /cgn2\_6/ptodata/2/pna/US6040\_COMB.seq:\*
- 97: /cgn2\_6/ptodata/2/pna/US6041\_COMB.seq:\*
- 98: /cgn2\_6/ptodata/2/pna/US6042\_COMB.seq:\*
- 99: /cgn2\_6/ptodata/2/pna/US6043\_COMB.seq:\*
- 100: /cgn2\_6/ptodata/2/pna/US6044\_COMB.seq:\*
- 101: /cgn2\_6/ptodata/2/pna/US6045\_COMB.seq:\*
- 102: /cgn2\_6/ptodata/2/pna/US6046\_COMB.seq:\*
- 103: /cgn2\_6/ptodata/2/pna/US6047\_COMB.seq:\*
- 104: /cgn2\_6/ptodata/2/pna/US6048\_COMB.seq:\*
- 105: /cgn2\_6/ptodata/2/pna/US6049\_COMB.seq:\*
- 106: /cgn2\_6/ptodata/2/pna/US6050\_COMB.seq:\*
- 107: /cgn2\_6/ptodata/2/pna/US6051\_COMB.seq:\*
- 108: /cgn2\_6/ptodata/2/pna/US6052\_COMB.seq:\*
- 109: /cgn2\_6/ptodata/2/pna/US6053\_COMB.seq:\*
- 110: /cgn2\_6/ptodata/2/pna/US6054\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1454.2	99.9	1455	1	PCT-US02-34674-3	Sequence 3, Appl1
2	1454.2	99.9	1455	1	PCT-US02-34764-3	Sequence 3, Appl1
3	1454.2	99.9	1455	43	US-10-026-140-3	Sequence 3, Appl1
4	1454.2	99.9	1991	1	PCT-US02-34674-1	Sequence 1, Appl1
5	1454.2	99.9	1991	1	PCT-US02-34764-1	Sequence 1, Appl1
6	1454.2	99.9	1991	43	US-10-026-140-1	Sequence 1, Appl1
7	598	41.1	735	22	US-09-533-559-7458	Sequence 7458, Appl1
8	598	41.1	735	53	US-10-653-047-7458	Sequence 7458, Appl1
9	345.4	23.7	1572	20	US-09-417-507-9340	Sequence 9340, Appl1
10	334.4	23.9	741	20	US-09-417-507-19268	Sequence 19268, Appl1
11	318.8	21.9	2918	50	US-09-404-550-15403	Sequence 15403, Appl1
12	317.8	21.8	1470	50	US-10-368-483-37092	Sequence 37092, Appl1
13	317.8	21.8	1470	93	US-60-336-039-37092	Sequence 37092, Appl1
14	307.6	21.1	2914	69	US-60-138-103-7477	Sequence 7477, Appl1
15	294.6	20.2	1434	20	US-09-404-550-19029	Sequence 19029, Appl1
16	276.6	19.0	2037	51	US-10-437-963-97407	Sequence 97407, Appl1
17	266.6	18.3	1933	22	US-09-533-559-7405	Sequence 7405, Appl1
18	266.6	18.3	1933	53	US-10-653-047-7405	Sequence 7405, Appl1
19	249	17.1	458	20	US-09-417-507-14756	Sequence 14756, Appl1
20	246	16.9	1404	36	US-09-887-2724-42064	Sequence 4204, Appl1
21	233.4	16.0	1688	47	US-10-219-999-10366	Sequence 10066, Appl1
22	233.4	16.0	1688	51	US-10-425-114-4348	Sequence 4348, Appl1
23	233.4	16.0	1688	51	US-10-425-114-4348	Sequence 4348, Appl1
24	231.8	15.9	1711	30	US-09-708-477-79294	Sequence 79294, Appl1
25	230.4	15.8	1794	47	US-10-219-999-14009	Sequence 14009, Appl1
26	230.4	15.8	1794	51	US-10-425-114-19747	Sequence 19747, Appl1
27	230.4	15.8	1794	51	US-10-425-114-19747	Sequence 19747, Appl1
28	230.2	15.8	1712	30	US-09-708-477-54992	Sequence 54992, Appl1
29	227.4	15.6	1781	47	US-10-219-999-29391	Sequence 29391, Appl1
30	227.4	15.6	1781	51	US-10-425-114-30939	Sequence 30939, Appl1
31	227.4	15.6	1781	51	US-10-425-114A-30939	Sequence 30939, Appl1
32	227.4	15.6	1931	47	US-10-219-999-29934	Sequence 29934, Appl1
33	227.4	15.6	1931	51	US-10-425-114-30835	Sequence 30835, Appl1
34	227.4	15.6	1931	51	US-10-425-114A-30835	Sequence 30835, Appl1
35	227.4	15.6	1932	47	US-10-219-999-16845	Sequence 16845, Appl1
36	227.4	15.6	1932	51	US-10-425-114-33175	Sequence 33175, Appl1
37	227.4	15.6	1932	51	US-10-425-114A-33175	Sequence 33175, Appl1
38	227.4	15.6	1992	27	US-09-654-617-262950	Sequence 262950, Appl1
39	227.4	15.6	1992	29	US-09-684-016-262950	Sequence 262950, Appl1
40	219.2	15.1	1737	47	US-10-219-999-15047	Sequence 15047, Appl1
41	219.2	15.1	1737	51	US-10-425-114-25892	Sequence 25892, Appl1
42	219.2	15.1	1737	51	US-10-425-114A-25892	Sequence 25892, Appl1
43	214.6	14.7	1503	36	US-09-687-27284-2847	Sequence 2847, Appl1
44	214.6	14.7	1814	51	US-10-437-963-92182	Sequence 92182, Appl1
45	206.8	14.2	1843	54	US-10-733-930-4492	Sequence 4492, Appl1

## ALIGNMENTS

```

RESULT 1
PCT-US02-34674-3
Sequence 3, Application PC/TUS0234674
GENERAL INFORMATION:
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Goedegebuurt, Frits
APPLICANT: Ward, Michael
APPLICANT: Yao, Jian
TITLE OF INVENTION: BG15 Beta-Glucosidase and Nucleic Acid
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: GC697
CURRENT APPLICATION NUMBER: PCT/US02/34674
CURRENT FILING DATE: 2003-03-14
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 1455
TYPE: DNA
ORGANISM: Trichoderma reesei

```

PCT-US02-34674-3									
Query Match	99.9%	Score 1454.2	DB 1	Length 1455					
Best Local Similarity	100.0%	Pred. No. 0							
Matches 1455	Conservative 0	Mismatches 0	Indels 0	Gaps 0					
QY	1	ATGCCCGAGTGCAGAGCTTGCCCAAGACTTTGAATGAGGCTTCGCAACGGCCGCTTAC	60						
DB	1	ATGCCCGAGTGCAGAGCTTGCCCAAGACTTTGAATGAGGCTTCGCAACGGCCGCTTAC	60						
QY	61	CAGATCGAAGCGCCGTCAAGAAAGTGGCCGCGCCCGTCCATCTGGGACACGTAATGC	120						
DB	61	CAGATCGAAGCGCCGTCAAGAAAGTGGCCGCGCCCGTCCATCTGGGACACGTAATGC	120						
QY	121	CACCTGGAAGCATTGGCGCACCAAGGGGCCAACGGCCGATGTGGCTTGGATCACTACAC	180						
DB	121	CACCTGGAAGCATTGGCGCACCAAGGGGCCAACGGCCGATGTGGCTTGGATCACTACAC	180						
QY	181	CGCTACGATGAGACTTTGATCTCTTACCAAGTACGGCGCAAAAGGCTTACCGCTTCC	240						
DB	181	CGCTACGATGAGACTTTGATCTCTTACCAAGTACGGCGCAAAAGGCTTACCGCTTCC	240						
QY	241	TTGTGATGTCGCGGATCATTTCCCTCGCGCGCAGCTGATCCCGTCAACGAGAGGGA	300						
DB	241	TTGTGATGTCGCGGATCATTTCCCTCGCGCGCAGCTGATCCCGTCAACGAGAGGGA	300						
QY	301	ATTGAGTTTTCACGAAACTGATTTGAGCGCTCTTTGAGCGGGGTATCACGCTTTGGTG	360						
DB	301	ATTGAGTTTTCACGAAACTGATTTGAGCGCTCTTTGAGCGGGGTATCACGCTTTGGTG	360						
QY	361	ACTTTGATCCATGAGGATCTGCGTCAAGCGCTTACAGATCGCTATGAGAGGTGGCTCAAC	420						
DB	361	ACTTTGATCCATGAGGATCTGCGTCAAGCGCTTACAGATCGCTATGAGAGGTGGCTCAAC	420						
QY	421	GTGGAAGAGGTCCAGCTGACTTTGAGCGGTATCGAGTGTGCTTTGAACGTTTTGGG	480						
DB	421	GTGGAAGAGGTCCAGCTGACTTTGAGCGGTATCGAGTGTGCTTTGAACGTTTTGGG	480						
QY	481	GACCGAGTCCAGAACTGGATACCATCAACAMCCTCGAATTCAGGCGCATTTAGAAAT	540						
DB	481	GACCGAGTCCAGAACTGGATACCATCAACAMCCTCGAATTCAGGCGCATTTAGAAAT	540						
QY	541	GGCACCGGACCAAGCCCGGGAGAGGACGACTTAAACAAGCACTCCACGAGGGCAAC	600						
DB	541	GGCACCGGACCAAGCCCGGGAGAGGACGACTTAAACAAGCACTCCACGAGGGCAAC	600						
QY	601	ACTCCACTGAGCCGTGGCTGCTGTAAGAAAGGCCAGATCATGAGCCATGCCGCGCTG	660						
DB	601	ACTCCACTGAGCCGTGGCTGCTGTAAGAAAGGCCAGATCATGAGCCATGCCGCGCTG	660						
QY	661	GCCGCTTACAGACAGGACTTTCGCCCTTCGCAAAAGGGCCAGATCGGATCTTCCTCAAC	720						
DB	661	GCCGCTTACAGACAGGACTTTCGCCCTTCGCAAAAGGGCCAGATCGGATCTTCCTCAAC	720						
QY	721	GGCGACTTATGAGCCCTGGGACAGCAATAGCCTCGGACCAAGAGAGCTGTGAGGGA	780						
DB	721	GGCGACTTATGAGCCCTGGGACAGCAATAGCCTCGGACCAAGAGAGCTGTGAGGGA	780						
QY	781	CGGATGGAATTTACATTTGGCTGGTTTCCCAATCCATCTTCTTGAAGAAAGACTATCA	840						
DB	781	CGGATGGAATTTACATTTGGCTGGTTTCCCAATCCATCTTCTTGAAGAAAGACTATCA	840						
QY	841	GAGAGCATGAAGAACAGCTGGGCGAGAGGCTTCCAGCCTCACTCCCGCGACTTTGGC	900						
DB	841	GAGAGCATGAAGAACAGCTGGGCGAGAGGCTTCCAGCCTCACTCCCGCGACTTTGGC	900						
QY	901	ATCTCTCAATGCCGAGAGACCGAATTTACGAGCATGAATTTACACATCCCAAGTTCCG	960						
DB	901	ATCTCTCAATGCCGAGAGACCGAATTTACGAGCATGAATTTACACATCCCAAGTTCCG	960						
QY	961	CGCCACCTGAGCGGTCCTCCGAGCGGACTTATCTTCGCGCCATCATGAGACCAAG	1020						
DB	961	CGCCACCTGAGCGGTCCTCCGAGCGGACTTATCTTCGCGCCATCATGAGACCAAG	1020						

Qy	1021	GAGAAATAAGACGGCAGCCCGCTGGCGAGGAGCGGCGCTCGCTGCTGAGGCTCCGC	1080
Dp	1021	GAGAAATAAGACGGCAGCCCGCTGGCGAGGAGCGGCGCTCGCTGCTGAGGCTCCGC	1080
Qy	1081	CCGGACATGTGTCGGGAAGCATCTGCCCGGGGTGTACCGGCTGTACCGGACAGCCCATCTAC	1140
Dp	1081	CCGGACATGTGTCGGGAAGCATCTGCCCGGGGTGTACCGGCTGTACCGGACAGCCCATCTAC	1140
Qy	1141	ATCACCCGAGAACGGATGCGCCGCTGCCCTTGGAAGAGAAATGACGTGCGAGAGGCGCTC	1200
Dp	1141	ATCACCCGAGAACGGATGCGCCGCTGCCCTTGGAAGAGAAATGACGTGCGAGAGGCGCTC	1200
Qy	1201	AACACACCCCTTCGCGATCCGCTACTTTGACTCGCATCTGAGCTGATTTCCAAAGCCATT	1260
Dp	1201	AACACACCCCTTCGCGATCCGCTACTTTGACTCGCATCTGAGCTGATTTCCAAAGCCATT	1260
Qy	1261	ACCCAGAGCGGCGTGTCTGTCAAGGGATCTTTTGGCTGGGCGTTGCTCATTACTTGGA	1320
Dp	1261	ACCCAGAGCGGCGTGTCTGTCAAGGGATCTTTTGGCTGGGCGTTGCTCATTACTTGGA	1320
Qy	1321	TGCTCAGATGCTTACGGACCCAGATTTGGGGCTCACGTTACAGACTACACCAACCTCAAG	1380
Dp	1321	TGCTCAGATGCTTACGGACCCAGATTTGGGGCTCACGTTACAGACTACACCAACCTCAAG	1380
Qy	1381	CGCAGCGCCCAAGAAAGTCTGCCCTGCTCTCAAGACATGTTTGGCGCCCGGCAAGGGTT	1440
Dp	1381	CGCAGCGCCCAAGAAAGTCTGCCCTGCTCTCAAGACATGTTTGGCGCCCGGCAAGGGTT	1440
Qy	1441	AAAGTGGCGGCATTA	1455
Dp	1441	AAAGTGGCGGCATTA	1455

```

RESULT 2
PCT-US02-34764-3
; Sequence 3, Application PC/TUS0234764
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuer, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: PCT/US02/34764
; CURRENT FILING DATE: 2003-10-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
; PCT-US02-34764-3

```

Query Match	99.9%	Score 1454.2	DB 1	Length 1455
Best Local Similarity	100.0%	Pred. NO. 0		
Matches 1455	0	Mismatches	0	Gaps 0

QY	1	ATGCCGAGTGTGCTAGCTCTGCCCCAGACCTTTGAAATGGGGCTTCGCAAGCGGCGCTAC	60
Db	1	ATGCCCGAGTGTGCTAGCTCTGCCCCAGACCTTTGAAATGGGGCTTCGCAAGCGGCGCTAC	60
QY	61	CAGATCGAAGGCGCGCTCAAGAAAGTGTCGCGCGCCGTCATCTGGGACAGCTATGC	120
Db	61	CAGATCGAAGGCGCGCTCAAGAAAGTGTCGCGCGCCGTCATCTGGGACAGCTATGC	120
QY	121	CACCTGGAGCATGCGCGCACCAACGGGCGCCAAACGGCGATGTGGCTTGGGATCACTAACAC	180
Db	121	CACCTGGAGCATGCGCGCACCAACGGGCGCCAAACGGCGATGTGGCTTGGGATCACTAACAC	180
QY	181	CGTACGATGAGGACTTTGATCTCTTGACCAAGTACGGCGCAAGAGCTTACCGCTTCTCC	240

Db	181	CGCTACGATGAGGACTTTGATCTCTTGACCAAGTACGGCGCAAGGGCTACCGCTTCC	240
OY	241	TTGTGCTGTCGCGGATCATTCCTCTCGCGGGCAGGCTGATCCCTGTACAGAGAGGA	300
Db	241	TTGTGCTGTCGCGGATCATTCCTCTCGCGGGCAGGCTGATCCCTGTACAGAGAGGA	300
OY	301	ATTGAGTTTTCACACAACTGATTGAGCCCTGTGTAAGCGGGGTATCACGCTTTGGGTG	360
Db	301	ATTGAGTTTTCACACAACTGATTGAGCCCTGTGTAAGCGGGGTATCACGCTTTGGGTG	360
OY	361	ACTTTGTACCACTGGGATCTGCTCAGAGCCTTACCATCGCTATGAGAGGTGCTCAAC	420
Db	361	ACTTTGTACCACTGGGATCTGCTCAGAGCCTTACCATCGCTATGAGAGGTGCTCAAC	420
OY	421	GTGGAAGAGGTCCAGCTTGTGAGCCGATGTCGAGGTTGTCTTTGAACGTTTTGGG	480
Db	421	GTGGAAGAGGTCCAGCTTGTGAGCCGATGTCGAGGTTGTCTTTGAACGTTTTGGG	480
OY	481	GACCGAGTCCGAACCTGGATATCACATCAACGACCCCTTGATTTCAAGGCCATTTATGATAT	540
Db	481	GACCGAGTCCGAACCTGGATATCACATCAACGACCCCTTGATTTCAAGGCCATTTATGATAT	540
OY	541	GCACACGGGACGAAACGCCCCGGGGAGAGACGACTTTAACAGCACTCCACCGAGGCAAC	600
Db	541	GCACACGGGACGAAACGCCCCGGGGAGAGACGACTTTAACAGCACTCCACCGAGGCAAC	600
OY	601	ACTGCCACTGAGCCGTGGCTGCTGTAAGAAAGGCCAGATCATGAGCCATGCCCCGCGTG	660
Db	601	ACTGCCACTGAGCCGTGGCTGCTGTAAGAAAGGCCAGATCATGAGCCATGCCCCGCGTG	660
OY	661	GCCGTCTACAGCAGAGGACTTTGCCCCCTCGCAAAAAGGCCAGATCGGCAATTCGCTCAAC	720
Db	661	GCCGTCTACAGCAGAGGACTTTGCCCCCTCGCAAAAAGGCCAGATCGGCAATTCGCTCAAC	720
OY	721	GGCGACTACTATGAGCCCTGGGAGCAGCAATGAGCTGGGAGCAAGAGGGCTGTGAGCGA	780
Db	721	GGCGACTACTATGAGCCCTGGGAGCAGCAATGAGCTGGGAGCAAGAGGGCTGTGAGCGA	780
OY	781	CGGATGGAATTTCAATTTGCTGTTTGGCCATCCATCTTCTTGAAGAGAGACTATCA	840
Db	781	CGGATGGAATTTCAATTTGCTGTTTGGCCATCCATCTTCTTGAAGAGAGACTATCA	840
OY	841	GAGAGCATGAAGAGCAGCTGGGCGAAGGCTTCAAGCCTCACTCCCGCGGACTTTGGCC	900
Db	841	GAGAGCATGAAGAGCAGCTGGGCGAAGGCTTCAAGCCTCACTCCCGCGGACTTTGGCC	900
OY	901	ATCTCTCAATGCGGAGAGACCGACTTTCACGCGATGAATTAATAACAATCCCAATTTGGCG	960
Db	901	ATCTCTCAATGCGGAGAGACCGACTTTCACGCGATGAATTAATAACAATCCCAATTTGGCG	960
OY	961	CGCCACCTAGACGAGTCCCGTCCCGAGACCGACTATCTCGGCGGCATTCATGACACACAG	1020
Db	961	CGCCACCTAGACGAGTCCCGTCCCGAGACCGACTATCTCGGCGGCATTCATGACACACAG	1020
OY	1021	GAGAAATAAGACGCGCAGCCCGTGTGGCGAGAGAGCGGCTCGCTGAGCTGCGTCTTGC	1080
Db	1021	GAGAAATAAGACGCGCAGCCCGTGTGGCGAGAGAGCGGCTCGCTGAGCTGCGTCTTGC	1080
OY	1081	CCGGACATATTTCCGGAGAGCATCTGCGCCCGGCTGTACCGGCTGTACCGGACACCATCTAC	1140
Db	1081	CCGGACATATTTCCGGAGAGCATCTGCGCCCGGCTGTACCGGCTGTACCGGACACCATCTAC	1140
OY	1141	ATCACCGAGAACGATGCGCGTGGCTCTGAGAGAGAGAAATGACGTGCGAGAGAGCGCTC	1200
Db	1141	ATCACCGAGAACGATGCGCGTGGCTCTGAGAGAGAGAAATGACGTGCGAGAGAGCGCTC	1200
OY	1201	AAGGACCCCTTTCGGCATTCGSAATCTTTGACTCGGCACTTGAATCGATTTTCCAGAGCCATT	1266
Db	1201	AAGGACCCCTTTCGGCATTCGSAATCTTTGACTCGGCACTTGAATCGATTTTCCAGAGCCATT	1266
OY	1261	ACCGAGAGCGGCGTCTGCTCAAGGGATATCTTTGCGTGGCGGCTTGTCTCGATTAATTGGAA	1320
Db	1261	ACCGAGAGCGGCGTCTGCTCAAGGGATATCTTTGCGTGGCGGCTTGTCTCGATTAATTGGAA	1320

```
QY 1321 TGTGATGATGCTACGAGCCGAGATTGGGCGTCAAGTTCAGAGACTACAGCACTCCCTGAG 1380
DB 1321 TGTGATGATGCTACGAGCCGAGATTGGGCGTCAAGTTCAGAGACTACAGCACTCCCTGAG 1380
QY 1381 CGCAGCGCCCAAGAACTGCTGCTGCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
DB 1381 CGCAGCGCCCAAGAACTGCTGCTGCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
QY 1441 AAAGTGGCGGCATTA 1455
DB 1441 AAAGTGGCGGCATTA 1455

RESULT 3
US-10-026-140-3
; Sequence 3, Application US/10026140
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-3

Query Match 99.9%; Score 1454.2; DB 43; Length 1455;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCGAGTGTGCTAGCTCTGCGCCCAAGCACTTTGATGGGCTTCGCAACGGCGGCTTAC 60
DB 1 ATGCCGAGTGTGCTAGCTCTGCGCCCAAGCACTTTGATGGGCTTCGCAACGGCGGCTTAC 60
QY 61 CAGATCGAAGGCGCGTCAAGAAAGTGGCGCGCGCTTCATCTGGACAACGTACTGC 120
DB 61 CAGATCGAAGGCGCGTCAAGAAAGTGGCGCGCGCTTCATCTGGACAACGTACTGC 120
QY 121 CACCTGAGGCGATGCGGCACCAACGGCGCAAGTGGCTTGGCTGCTCACTACAC 180
DB 121 CACCTGAGGCGATGCGGCACCAACGGCGCAAGTGGCTTGGCTGCTCACTACAC 180
QY 181 CGCTACGATGAGCACTTTGATCTCTTGACCAAGTACGCGCAAGGCTTACCGCTTCC 240
DB 181 CGCTACGATGAGCACTTTGATCTCTTGACCAAGTACGCGCAAGGCTTACCGCTTCC 240
QY 241 TTTGCTGCTGCTGCGGATCATTTCCCTTGCGGCGAGGCTGATCCGTCACAGAGAGGA 300
DB 241 TTTGCTGCTGCTGCGGATCATTTCCCTTGCGGCGAGGCTGATCCGTCACAGAGAGGA 300
QY 301 ATTGAGTTTACAGCAAACTGATGAGCGCTGTGAGGGGGATACAGGCTTGGGTG 360
DB 301 ATTGAGTTTACAGCAAACTGATGAGCGCTGTGAGGGGGATACAGGCTTGGGTG 360
QY 361 ACTTTGATCACTGAGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTAAC 420
DB 361 ACTTTGATCACTGAGATCTGCTCAGGCGCTTCAAGATGCTATGAGAGGCTGCTAAC 420
QY 421 GTGGAAGAGGTCCAGCTGAGCACTTTGAGCGGTATGCGAGTGTGCTTTGAAGCTTTGGG 480
DB 421 GTGGAAGAGGTCCAGCTGAGCACTTTGAGCGGTATGCGAGTGTGCTTTGAAGCTTTGGG 480
QY 481 GACCGAGTCCAGAACTGATCAACATCAAGWCCGTGATTGAGGCACTATGATAT 540
DB 481 GACCGAGTCCAGAACTGATCAACATCAAGWCCGTGATTGAGGCACTATGATAT 540
```

```
DB 481 GACCGAGTCCAGAACTGATCAACATCAAGWCCGTGATTGAGGCACTATGATAT 540
QY 541 GCCACCGGACAGCAACGCGCCCGGACAGAGAGCATTTAACAAGACTTCACCGAGGCAAC 600
DB 541 GCCACCGGACAGCAACGCGCCCGGACAGAGAGCATTTAACAAGACTTCACCGAGGCAAC 600
QY 601 ACTGCCACTGAGCGCTGCTGCTGGAAGAGCCCAAGTCAATGAGCCATGCTCCGCGCTG 660
DB 601 ACTGCCACTGAGCGCTGCTGCTGGAAGAGCCCAAGTCAATGAGCCATGCTCCGCGCTG 660
QY 661 GCCGTCTACAGCAGGACTTTGCGCCCTCGCAAAAGGCGCAGATCCGCACTTCGCTAAC 720
DB 661 GCCGTCTACAGCAGGACTTTGCGCCCTCGCAAAAGGCGCAGATCCGCACTTCGCTAAC 720
QY 721 GCGGACTACTATGAGCCCTGAGGACAGCATGAGCTTCGAGGACAAAGAGCTGCTGAGCGA 780
DB 721 GCGGACTACTATGAGCCCTGAGGACAGCATGAGCTTCGAGGACAAAGAGCTGCTGAGCGA 780
QY 781 CGGATGGAAATTTCAATTTGGCTGTTGGCAATCCCATCTTTGAAAGAGCTATCCCA 840
DB 781 CGGATGGAAATTTCAATTTGGCTGTTGGCAATCCCATCTTTGAAAGAGCTATCCCA 840
QY 841 GAGAGCATGAAGAAGAGCTGGGCGAGAGCTTCCAGCCCTCACTCCCGGACTTTGGCC 900
DB 841 GAGAGCATGAAGAAGAGCTGGGCGAGAGCTTCCAGCCCTCACTCCCGGACTTTGGCC 900
QY 901 ATCCTTAAGCGCGAAGAACCCGACTTCTACGGATGAATTAATCAATCCCATGCTGGG 960
DB 901 ATCCTTAAGCGCGAAGAACCCGACTTCTACGGATGAATTAATCAATCCCATGCTGGG 960
QY 961 CGCCACTTACAGCGGTCCGCTCCCGAGACGACTATCTGCGCGCCATCCATGAGACACAG 1020
DB 961 CGCCACTTACAGCGGTCCGCTCCCGAGACGACTATCTGCGCGCCATCCATGAGACACAG 1020
QY 1021 GAGAATTAAGAGCGGACGCCCTGTTGGCGAGAGAGCGGCTTGGCTGCTGCTTGC 1080
DB 1021 GAGAATTAAGAGCGGACGCCCTGTTGGCGAGAGAGCGGCTTGGCTGCTGCTTGC 1080
QY 1081 CGGACATGTTCCGGAAGCATCTGCGCGGGGTACAGGCTGTGACGCGCAAGCCATCTAC 1140
DB 1081 CGGACATGTTCCGGAAGCATCTGCGCGGGGTACAGGCTGTGACGCGCAAGCCATCTAC 1140
QY 1141 ATCACCAGAAAGGATGCCGCTGCTGAGAGAGAAACATGACGTCGAGAGAGGCGCTC 1200
DB 1141 ATCACCAGAAAGGATGCCGCTGCTGAGAGAGAAACATGACGTCGAGAGAGGCGCTC 1200
QY 1201 AAGAACCCCTTCCGATCCGCTACTTGTGACTGCTGCACTTGAATTCGAAGGCATT 1260
DB 1201 AAGAACCCCTTCCGATCCGCTACTTGTGACTGCTGCACTTGAATTCGAAGGCATT 1260
QY 1261 ACCCAGAGCGGCGTGTGCTCAAGGGGTACTTTGGGTGGGCGTGTGCTGATTAATTGGA 1320
DB 1261 ACCCAGAGCGGCGTGTGCTCAAGGGGTACTTTGGGTGGGCGTGTGCTGATTAATTGGA 1320
QY 1321 TGTGATGATGCTACGAGCCAGATTCGCGCTCAAGTTCACAGACTACCAACCTCAAG 1380
DB 1321 TGTGATGATGCTACGAGCCAGATTCGCGCTCAAGTTCACAGACTACCAACCTCAAG 1380
QY 1381 CGCAGCGCCCAAGAACTGCTGCTGCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
DB 1381 CGCAGCGCCCAAGAACTGCTGCTGCTCTCAAGACATGTTGGCGCCCGCAGAGGTT 1440
QY 1441 AAAGTGGCGGCATTA 1455
DB 1441 AAAGTGGCGGCATTA 1455

RESULT 4
PCT-US02-34674-1
; Sequence 1, Application PC/TUS0234674
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
```

; APPLICANT: Ward, Michael  
 ; APPLICANT: Yao, Jian  
 ; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids  
 ; TITLE OF INVENTION: Encoding the Same  
 ; FILE REFERENCE: GC697  
 ; CURRENT APPLICATION NUMBER: PCT/US02/34674  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1991  
 ; TYPE: DNA  
 ; ORGANISM: Trichoderma reesei  
 ; PCT-US02-34674-1

Query Match 99.9%; Score 1454.2; DB 1; Length 1991;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCCGAGTCGCTAGCTGTGCCCAAGACTTTGATGGGGCTTCGCAACGGCGCCCTAC 60  
 DB 62 ATGCCCGAGTCGCTAGCTGTGCCCAAGACTTTGATGGGGCTTCGCAACGGCGCCCTAC 121  
 QY 61 CAGATCGAAGCGCCGTCGCAAGAGGTGGCGCGCCGCTTCATCTGAGACACGCTACTGC 120  
 DB 122 CAGATCGAAGCGCCGTCGCAAGAGGTGGCGCGCCGCTTCATCTGAGACACGCTACTGC 181  
 QY 121 CAGCTGAGCGCTTCGCGCAACGCGCGCAACGCGCGATGGCTTGGCATCTTACAC 180  
 DB 182 CAGCTGAGCGCTTCGCGCAACGCGCGCAACGCGCGATGGCTTGGCATCTTACAC 241  
 QY 181 CGCTACGATGAGACTTGTGATCTCTTGAACCAAGTACGCGCAAGGCGCTACCGCTTCTCC 240  
 DB 242 CGCTACGATGAGACTTGTGATCTCTTGAACCAAGTACGCGCAAGGCGCTACCGCTTCTCC 301  
 QY 241 TTGTGCTGTCGCGGATCATTTCCCTCGCGCGAGGCTGTATCCCGTCAACGAGAGGGA 300  
 DB 302 TTGTGCTGTCGCGGATCATTTCCCTCGCGCGAGGCTGTATCCCGTCAACGAGAGGGA 361  
 QY 301 ATTGAGCTTTTACAGCAAACTGATTTGAGCGCTTGTGAGGGGGGTATCAAGCTTTGGGTG 360  
 DB 362 ATTGAGCTTTTACAGCAAACTGATTTGAGCGCTTGTGAGGGGGGTATCAAGCTTTGGGTG 421  
 QY 361 ACTTTGACCACTGGGATCTGCTCAGCGCGCTTACAGATGCTATGAGGGCTGCTCAAC 420  
 DB 422 ACTTTGACCACTGGGATCTGCTCAGCGCGCTTACAGATGCTATGAGGGCTGCTCAAC 481  
 QY 421 GTGGAAGAGTCCAGCTTGAATTTGAGCGGTATCGAGGTTGTCTTTGAACTTTTGGG 480  
 DB 482 GTGGAAGAGTCCAGCTTGAATTTGAGCGGTATCGAGGTTGTCTTTGAACTTTTGGG 541  
 QY 481 GACCGAGTCCAGAACTGATATCAATCAACGAGCCCTTGATTCAGCGCATTTATGATAT 540  
 DB 542 GACCGAGTCCAGAACTGATATCAATCAACGAGCCCTTGATTCAGCGCATTTATGATAT 601  
 QY 541 GCCACGGGAGCAAGCGCCGGGAGAGAGCATTTAAACAAGCACTCAACGAGAGGGAAC 600  
 DB 602 GCCACGGGAGCAAGCGCCGGGAGAGAGCATTTAAACAAGCACTCAACGAGAGGGAAC 661  
 QY 601 ACTGCCACTGAGCGCTGCTGTGAAAGGCCCAAGATCATAGACCATGCCCCGGCGGTG 660  
 DB 662 ACTGCCACTGAGCGCTGCTGTGAAAGGCCCAAGATCATAGACCATGCCCCGGCGGTG 721  
 QY 661 GCCGCTTACAGAGGACTTTCCGCCCTCGCAAAAAGGCCCAAGATCGCGCATTCCTCTCAAC 720  
 DB 722 GCCGCTTACAGAGGACTTTCCGCCCTCGCAAAAAGGCCCAAGATCGCGCATTCCTCTCAAC 781  
 QY 721 GGGGACTACTATGAGCGCTGGGAGCAGCAATGAGGCTGGGAGCAAGAGGCTGTGAGGGA 780  
 DB 782 GGGGACTACTATGAGCGCTGGGAGCAGCAATGAGGCTGGGAGCAAGAGGCTGTGAGGGA 841  
 QY 781 CGGATGGAATTTCACTATGGCTGTTGGCAATCCATCTTTTGAAGAAGACTATATCA 840

DB 842 CGGATGGAATTTCACTATGGCTGTTGGCAATCCATCTTTTGAAGAAGACTATATCA 901  
 QY 841 GAGAGCATGAAGAAGAGCTGGGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTTGGC 900  
 DB 902 GAGAGCATGAAGAAGAGCTGGGAGAGGCTTCAAGCCCTCACTCCCGGAGCTTTGGC 961  
 QY 901 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAATCCAGTTGGCG 960  
 DB 962 ATCTCAATGCGGAGAGACCGACTTCTACGCGATGATTAACAATCCAGTTGGCG 1021  
 QY 961 CGCCACTTAAAGCGTCCGCTCCCGGAGACGCACTATTTGGCGCGCATCTCAAGACACAG 1020  
 DB 1022 CGCCACTTAAAGCGTCCGCTCCCGGAGACGCACTATTTGGCGCGCATCTCAAGACACAG 1081  
 QY 1021 GAGATTAAGAGCGGACGCGCGCTTGGGAGAGAGGCGCTCGCTGGCTGCTCTGC 1080  
 DB 1082 GAGATTAAGAGCGGACGCGCGCTTGGGAGAGAGGCGCTCGCTGGCTGCTCTGC 1141  
 QY 1081 CCGGACATGTTCCGGAAGCATCTTCCCGGCTGTACGCGCTGTACGCGCAAGCCCATCTAC 1140  
 DB 1142 CCGGACATGTTCCGGAAGCATCTTCCCGGCTGTACGCGCTGTACGCGCAAGCCCATCTAC 1201  
 QY 1141 ATCAACGAGAAAGGATGCGCGCTGCTGAGAGAGAAATGACGTCGAGAGGCCGCTC 1200  
 DB 1202 ATCAACGAGAAAGGATGCGCGCTGCTGAGAGAGAAATGACGTCGAGAGGCCGCTC 1261  
 QY 1201 AAGGACCCCTTCCGCAATCGCTCTTGTGATCCGCACTTGGACTTGCATTTCCAGGCGCAT 1260  
 DB 1262 AAGGACCCCTTCCGCAATCGCTCTTGTGATCCGCACTTGGACTTGCATTTCCAGGCGCAT 1321  
 QY 1261 ACCGAGACCGGCTGCTGCTCAAGGGGTACTTTGCGTGGCGCTGTGATTAACCTTGGAA 1320  
 DB 1322 ACCGAGACCGGCTGCTGCTCAAGGGGTACTTTGCGTGGCGCTGTGATTAACCTTGGAA 1381  
 QY 1321 TGGTCAGATGCTTACGGAACCAATTTGCGGTGATCTTCAACAGCTTACACCCCTTAAG 1380  
 DB 1382 TGGTCAGATGCTTACGGAACCAATTTGCGGTGATCTTCAACAGCTTACACCCCTTAAG 1441  
 QY 1381 CGGACGCGCAAGAGTGGCGCTGCTCAAGAGCAATGTTTGGCGCGGAGAGGGTT 1440  
 DB 1442 CGGACGCGCAAGAGTGGCGCTGCTCAAGAGCAATGTTTGGCGCGGAGAGGGTT 1501  
 QY 1441 AAAGTGGCGGCAATTA 1455  
 DB 1502 AAAGTGGCGGCAATTA 1516

RESULT 5  
 PCT-US02-34764-1  
 ; Sequence 1, Application PC/TUS0234764  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Goedegebuur, Frits  
 ; APPLICANT: Ward, Michael  
 ; APPLICANT: Yao, Jian  
 ; TITLE OF INVENTION: BGL5 Beta-Glucosidase and Nucleic Acids  
 ; TITLE OF INVENTION: Encoding the Same  
 ; FILE REFERENCE: GC697  
 ; CURRENT APPLICATION NUMBER: PCT/US02/34764  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 1991  
 ; TYPE: DNA  
 ; ORGANISM: Trichoderma reesei  
 ; PCT-US02-34764-1

Query Match 99.9%; Score 1454.2; DB 1; Length 1991;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGCCCGAGTCGCTAGCTGTGCCCAAGACTTTGATGGGGCTTCGCAACGGCGCCCTAC 60

```
Db 62 ATGCCGAGTCCGCTAGCTCTGCCCCAAGCATTGATGGGCTTCGCAACGGCCGCTAC 121
Qy 61 CAGATCGAAGGCGCGGTCAAGAAGTGGCGCGGCGCGTCCATCTGGGACACCTATCC 120
Db 122 CAGATCGAAGGCGCGGTCAAGAAGTGGCGCGGCGCGTCCATCTGGGACACCTATCC 181
Qy 121 CACCTGAGGCGCGGTCAAGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
Db 182 CACCTGAGGCGCGGTCAAGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 241
Qy 181 CGCTACGATGAGGACTTGTATCTTTGACCAAGTACGGGCGCAAGGCGGCGGCGGCGG 240
Db 242 CGCTACGATGAGGACTTGTATCTTTGACCAAGTACGGGCGCAAGGCGGCGGCGGCGG 301
Qy 241 TTGTGCGGTGCGGATCAATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 302 TTGTGCGGTGCGGATCAATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361
Qy 301 ATTGAGTTTTCAGCAAACTGATTTGACGCGCTGTTGAGGCGGCGGCGGCGGCGGCGG 360
Db 362 ATTGAGTTTTCAGCAAACTGATTTGACGCGCTGTTGAGGCGGCGGCGGCGGCGGCGG 421
Qy 361 ACTTGTACCACTGGGATCTGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 422 ACTTGTACCACTGGGATCTGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 481
Qy 421 GTGGAAGAGTTCAGCTGGAATTTGAGCGGATGAGGAGTGTGCTTGAAGCTTTGGG 480
Db 482 GTGGAAGAGTTCAGCTGGAATTTGAGCGGATGAGGAGTGTGCTTGAAGCTTTGGG 541
Qy 481 GACCGAGTTCAGCACTGATATCAATCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 540
Db 542 GACCGAGTTCAGCACTGATATCAATCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGG 601
Qy 541 GCGCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
Db 602 GCGCAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 661
Qy 601 ACTGCGACTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 662 ACTGCGACTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 721
Qy 661 GCGGCTTACAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 722 GCGGCTTACAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 781
Qy 721 GCGGCTTACAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 782 GCGGCTTACAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841
Qy 781 GCGATGGAATTTCACTTGGCTGTTTGCATTCCTTCTTGAAGAGGAGGAGGAGGAGG 840
Db 842 GCGATGGAATTTCACTTGGCTGTTTGCATTCCTTCTTGAAGAGGAGGAGGAGGAGG 901
Qy 841 GAGAGCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
Db 902 GAGAGCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 961
Qy 901 ATTCCTCATGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
Db 962 ATTCCTCATGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1021
Qy 961 CGCCACTTGAAGCGGCTCCGCGCGGAGCGGAGCTATCTCGGCGGCGGCGGCGGCGGCGG 1020
Db 1022 CGCCACTTGAAGCGGCTCCGCGCGGAGCGGAGCTATCTCGGCGGCGGCGGCGGCGGCGG 1081
Qy 1021 GAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
Db 1082 GAGAGTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1141
Qy 1081 CGGAGCATGTTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
```

```
Db 1142 CGGACATGTTCCGGAAGCATCTCCCGGGGTGTAAGGCGCTGTACGCGCAAGCCCATCTAC 1201
Qy 1141 ATCAACGAGAAAGGATCCCGTCCCTGGAGAGGAAACATGACGTGCGGAGGAGGCGGTC 1200
Db 1202 ATCAACGAGAAAGGATCCCGTCCCTGGAGAGGAAACATGACGTGCGGAGGAGGCGGTC 1261
Qy 1201 AACGACCCCTTCCGATCCGATCTTGTGATCTGCACTTGTGATCTGCACTTGTGATCTGCA 1260
Db 1262 AACGACCCCTTCCGATCCGATCTTGTGATCTGCACTTGTGATCTGCACTTGTGATCTGCA 1321
Qy 1261 ACCGAGAGCGGCGTCTGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
Db 1322 ACCGAGAGCGGCGTCTGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1381
Qy 1321 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
Db 1382 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441
Qy 1381 CGCAGCGCCAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1442 CGCAGCGCCAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1501
Qy 1441 AAAGTGGCGGATTA 1455
Db 1502 AAAGTGGCGGATTA 1516
```

## RESULT 6

```
US-10-026-140-1
; Sequence 1, Application US/10026140
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OR INVENTION: BGLS Beta-Glucosidase and Nucleic Acids
; FILE REFERENCE: GC697
; CURRENT APPLICATION NUMBER: US/10/026,140
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ. ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 1
; LENGTH: 1991
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-140-1
```

```
Query Match 99.9%; Score 1454.2; DB 43; Length 1991;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 ATGCCGAGTCCGCTAGCTCTGCCCCAAGCATTGATGGGCTTCGCAACGGCCGCTAC 60
Db 62 ATGCCGAGTCCGCTAGCTCTGCCCCAAGCATTGATGGGCTTCGCAACGGCCGCTAC 121
Qy 61 CAGATCGAAGGCGCGGTCAAGAAGTGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 122 CAGATCGAAGGCGCGGTCAAGAAGTGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 181
Qy 121 CACCTGAGGCGCGGTCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 180
Db 182 CACCTGAGGCGCGGTCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 241
Qy 181 CGCTACGATGAGGACTTGTATCTTTGACCAAGTACGGGCGCAAGGCGGCGGCGGCGGCGG 240
Db 242 CGCTACGATGAGGACTTGTATCTTTGACCAAGTACGGGCGCAAGGCGGCGGCGGCGGCGG 301
Qy 241 TTGTGCGGTGCGGATCAATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 302 TTGTGCGGTGCGGATCAATTCCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 361
Qy 301 ATTGAGTTTTCAGCAAACTGATTTGACGCGCTGTTGAGGCGGCGGCGGCGGCGGCGGCGG 360
```

|||||  
Db 362 ATTGAGTTTAAAGCAAACTGATTTGAGCGCCCTGTTGAGCGGGGATATACGCCCTTGGGATG 421  
Qy 361 ACTTTGACACATGGGATCTGCGCTCAGCGCTTACAGATGCTATGGAGGCTGGCTAAC 420  
Db 422 ACTTTGACACATGGGATCTGCGCTCAGCGCTTACAGATGCTATGGAGGCTGGCTAAC 481  
Qy 421 GTGGAAGAGGTCCAGCTGATCTTTGAGCGGATGCGAGGTTGTGCTTTGAAAGCTTTGGG 480  
Db 482 GTGGAAGAGGTCCAGCTGATCTTTGAGCGGATGCGAGGTTGTGCTTTGAAAGCTTTGGG 541  
Qy 481 GACCGAGTCCAGAACTGATATCATTCAAGAGCCCTGATTCAGGCGCATCTATGATAT 540  
Db 542 GACCGAGTCCAGAACTGATATCATTCAAGAGCCCTGATTCAGGCGCATCTATGATAT 601  
Qy 541 GCGACCGGAGCAAGCGCCCGGAGAGAGAGCATTAACAAGCATCCACAGAGGGCAAC 600  
Db 602 GCGACCGGAGCAAGCGCCCGGAGAGAGAGCATTAACAAGCATCCACAGAGGGCAAC 661  
Qy 601 ACTGCCACTGAGCGCTGCTCGCTGAAAGGCCAGATCATGAGCCATGCGCGCGCTG 660  
Db 662 ACTGCCACTGAGCGCTGCTCGCTGAAAGGCCAGATCATGAGCCATGCGCGCGCTG 721  
Qy 661 GCCGTCTACAGAGGGAATTTCCGCTTCCGAAAAGGCCAGATCGGCATCTGCTCAAC 720  
Db 722 GCCGTCTACAGAGGGAATTTCCGCTTCCGAAAAGGCCAGATCGGCATCTGCTCAAC 781  
Qy 721 GCGGACATCATAGAGCCCTGGAGACAGAAATGAGCTTGGAGCAAGAGAGCTTCTGAGACGA 780  
Db 782 GCGGACATCATAGAGCCCTGGAGACAGAAATGAGCTTGGAGCAAGAGAGCTTCTGAGACGA 841  
Qy 781 CGGATGGAATTTTCATTTGCTGCTGTTTGCATATCCATCTTTCTTGAAGAAGATATCCA 840  
Db 842 CGGATGGAATTTTCATTTGCTGCTGTTTGCATATCCATCTTTCTTGAAGAAGATATCCA 901  
Qy 841 GAGAGCATGAAGAAGCAAGCTTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCC 900  
Db 902 GAGAGCATGAAGAAGCAAGCTTGGGCGAGAGGCTTCCAGCCCTCACTCCCGGAGCTTTGCC 961  
Qy 901 ATCTCTCAATGCCGAGAGACCGAATTTAGGAGCATGAATTAATCAATCCCAATGCGG 960  
Db 962 ATCTCTCAATGCCGAGAGACCGAATTTAGGAGCATGAATTAATCAATCCCAATGCGG 1021  
Qy 961 CGCCACCTTACAGCGTCCGCTCCCGAGAGAGCACTATCTCGGCGCATCTCAAGACACAG 1020  
Db 1022 CGCCACCTTACAGCGTCCGCTCCCGAGAGAGCACTATCTCGGCGCATCTCAAGACACAG 1081  
Qy 1021 GAGAAATTAAGAGCGGACGCGCTGTTGGGAGAGAGCGGCTCGCTGAGCTGCTGTC 1080  
Db 1082 GAGAAATTAAGAGCGGACGCGCTGTTGGGAGAGAGCGGCTCGCTGAGCTGCTGTC 1141  
Qy 1081 CGGAGCATGTTTCCGGAAGCATCTCGCCCGGAGTGAACGCGCTGTAACGCGCAAGCCATCTAC 1140  
Db 1142 CGGAGCATGTTTCCGGAAGCATCTCGCCCGGAGTGAACGCGCTGTAACGCGCAAGCCATCTAC 1201  
Qy 1141 ATACACGAGAAGGAGTCCCGTGCCTGAGAGAGAGAACTGAGCTGCGAGAGGCGCTC 1200  
Db 1202 ATACACGAGAAGGAGTCCCGTGCCTGAGAGAGAGAACTGAGCTGCGAGAGGCGCTC 1261  
Qy 1201 AACGACCCCTTCCGATCCGATCTTGAATCGCATCTTGAATCTGATTTTCAAGGCCAT 1260  
Db 1262 AACGACCCCTTCCGATCCGATCTTGAATCGCATCTTGAATCTGATTTTCAAGGCCAT 1321  
Qy 1261 ACCGAGAGCGGCGCTGCTGTCAGAGGGGTAATTTGCGTGGCGCTTGCATATCTTGGAA 1320  
Db 1322 ACCGAGAGCGGCGCTGCTGTCAGAGGGGTAATTTGCGTGGCGCTTGCATATCTTGGAA 1381  
Qy 1321 TGGTCAGATGGCTACGGAACCAATTCGCGGCTCACTTCAAGACTACCAACCTTCAAG 1380  
Db 1382 TGGTCAGATGGCTACGGAACCAATTCGCGGCTCACTTCAAGACTACCAACCTTCAAG 1441  
Qy 1381 CGGACGCGCAAGAGTCTGCTGCTGCTCAAGGAGCATGTTTGGGCGCGGACAGAGGTT 1440  
|||||

Db 1442 CGGACGCGCAAGAGTCTGCGCTGCTCTTCAAGAGCATGTTTGGCGCGGACAGGGTT 1501  
Qy 1441 AAAGTGCGGCGCATTA 1455  
Db 1502 AAAGTGCGGCGCATTA 1516  
RESULT 7  
US-09-533-559-7458  
Sequence 7458, Application US/09533559  
GENERAL INFORMATION:  
APPLICANT: Randy M. Berka  
APPLICANT: Michael W. Rey  
APPLICANT: Jeffrey R. Shuster  
APPLICANT: Sakari Kauppinen  
APPLICANT: Ib Groth Clausen  
APPLICANT: Peter Bjarke Olsen  
TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
FILE REFERENCE: 5849.200-US  
CURRENT APPLICATION NUMBER: US/09/533,559  
EARLIER FILING DATE: 2000-03-22  
EARLIER APPLICATION NUMBER: 09/273,623  
NUMBER OF SEQ ID NOS: 7860  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7458  
LENGTH: 735  
TYPE: DNA  
ORGANISM: Tricoderma reesei  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) .. (735)  
OTHER INFORMATION: n = A,T,C or G  
US-09-533-559-7458  
Query Match 41.1%; Score 598; DB 22; Length 735;  
Best Local Similarity 97.0%; Pred. No. 1,4e-126;  
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;  
Qy 258 CATTCCCTTGGCGGAGGCTGATCCGTCACAGAGAGGAAATGAGTTTAAACAGCA 317  
Db 1 CATTCCCTTGGCGGAGGCTGATCCGTCACAGAGAGGAAATGAGTTTAAACAGCA 60  
Qy 318 ACTGATTGACGCGCTTGAAGCGGGTATCAAGCCCTGGGTGATCTTGTATCACTGGGGA 377  
Db 61 ACTGATTGACGCGCTTGAAGCGGGTATCAAGCCCTGGGTGATCTTGTATCACTGGGGA 120  
Qy 378 TCTGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCT 437  
Db 121 TCTGCTCAGGCGCTTACAGATGCTATGAGAGGCTGCTCAAGTGAAGAGGTCCAGCT 180  
Qy 438 GGAATTTGACGCGTATGCGAGTGTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTG 497  
Db 181 GGAATTTGACGCGTATGCGAGTGTGCTTTGAAGCTTTTGGGACCGAGTCCAGAACTG 240  
Qy 498 GATCACCATCAAGAGCCGCTGATTCAGGCGCATCTATGATATGCCACCGGAGCAAGCG 557  
Db 241 GATCACCATCAAGAGCCGCTGATTCAGGCGCATCTATGATATGCCACCGGAGCAAGCG 300  
Qy 558 CCGGAGAGAGAGATTAACAAGCACTCAACGAGGAGCAACTGCACTGAGCCGTG 617  
Db 301 CCGGAGAGAGAGATTAACAAGCACTCAACGAGGAGCAACTGCACTGAGCCGTG 360  
Qy 618 GTTGGCTGAAAGGCCAGATCATGAGCAATGCGCGCGCTGCTTCAACAGAGGGA 677  
Db 361 GTTGGCTGAAAGGCCAGATCATGAGCAATGCGCGCGCTGCTTCAACAGAGGGA 420  
Qy 678 CTTTGGCGCTTGCAGAAAGGCGAGATCGGCATCTGCTCAAGGAGGAGCACTATAGCC 737  
Db 421 CTTTGGCGCTTGCAGAAAGGCGAGATCGGCATCTGCTCAAGGAGGAGCACTATAGCC 480  
Qy 738 CTGGGACAGCAATGAGCTTCGGAGCAAGAGAGCTGCTGAGCGACGATGGAATTTCAAT 797  
|||||



```

Db      481  |||||CTGGAGAGCAATAGAGCTCGGGACAAAGAGGCTGAGAGGAGATGAAATTTCAAT 540
Qy      798  |||||TGGCTGTTGGCAATCCCATCTTCTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 856
Db      541  |||||TGGCTGTTGGCAATCCCATCTTCTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 600
Qy      857  AGCT--GGGCGAGAGGCTTCAGGCTTCAC-TCCGCGGACTTT--GCCATCTCAATGC 911
Db      601  AACTTGGGGGCGAGAGGCTTCAACCTTACTTCCGCGGACTTTCGCACTCTTAAATGC 660
Qy      912  CGGAGAGAC 921
Db      661  CGGAGAGAC 670

```

## RESULT 8

```

US-10-653-047-7458
; Sequence 7458, Application US/10653047
; GENERAL INFORMATION:
; APPLICANT: Randy M. Berka
; APPLICANT: Michael W. Rey
; APPLICANT: Jeffrey R. Shuster
; APPLICANT: Sakari Kauppinen
; APPLICANT: Ib Groth Clausen
; APPLICANT: Peter Bjørke Olsen
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 5849.200-US
; CURRENT APPLICATION NUMBER: US/10/653,047
; CURRENT FILING DATE: 2003-08-29
; PRIOR APPLICATION NUMBER: US/09/533,559
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/273,623
; PRIOR FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 7860
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7458
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(735)
; OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7458

```

```

Query Match      41.1%; Score 598; DB 53; Length 735;
Best Local Similarity 97.0%; Pred. No. 1.4e-126;
Matches 650; Conservative 1; Mismatches 13; Indels 6; Gaps 4;

```

```

Qy      258  CATTCCTCCCTGCGGAGAGCTGATCCCTCAACGAGAGGAGGAAATTTAGATTTACAGCA 317
Db      1     CATTCCTCCCTGCGGAGAGCTGATCCCTCAACGAGAGGAGGAAATTTAGATTTACAGCA 60
Qy      318  ACTGATTGACGCCCTGTGAGGCGGGATATCAAGCCTTGGATCTTTGATCCACTGGGA 377
Db      61  ACTGATTGACGCCCTGTGAGGCGGGATATCAAGCCTTGGATCTTTGATCCACTGGGA 120
Qy      378  TCTGCTCAGGCGCTTACAGATCGCTATGAGAGGCTGCTCAACGTGGAAGAGGTCCAGCT 437
Db      121  TCTGCTCAGGCGCTTACAGATCGCTATGAGAGGCTGCTCAACGTGGAAGAGGTCCAGCT 180
Qy      438  GGACTTTGAGCGGTATGCGGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTG 497
Db      181  GGACTTTGAGCGGTATGCGGAGGTGTGCTTTGAACGTTTGGGAGCCGAGTCCAGAACTG 240
Qy      498  GATCAACATCAAGAMCCTGTGATTCAAGGCCATCTATGATATGCCACCGGAGCAACGC 557
Db      241  GATCAACATCAAGAACCTGTGATTCAAGGCCATCTATGATATGCCACCGGAGCAACGC 300
Qy      558  CCGGCGGAGAGAGCATTTAACAAGACTCCACCGAGGGGAAACATCTGCCATGAGCCGTG 617

```

```

Db      301  CCGGCGGAGAGAGCATTTAACAAGACTCCACCGAGGGCAACACTGCCACTGAGCCGTG 360
Qy      618  GCTTCGTGGAAGAGCCCAATATGAGCCATGCCGCGCGCGGCGCTGTACAGAGGA 677
Db      361  GCTTCGTGGAAGAGCCCAATATGAGCCATGCCGCGCGCGGCGCTGTACAGAGGA 420
Qy      678  CTTTGCCTCCCGCAAAAAGGCGAGATCGGACTTTCGCTCAACGGGCACTACTATGAGCC 737
Db      421  CTTTGCCTCCCGCAAAAAGGCGAGATCGGACTTTCGCTCAACGGGCACTACTATGAGCC 480
Qy      738  CTGGAACAGCAATGAGCCTTCGGGACAGAGAGCTGTGAGCGACCGATGGAATTTCAAT 797
Db      481  CTGGAACAGCAATGAGCCTTCGGGACAGAGAGCTGTGAGCGACCGATGGAATTTCAAT 540
Qy      798  TGGCTGTTGGCAATCCCATCTTCTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 856
Db      541  TGGCTGTTGGCAATCCCATCTTCTGAAGAAGACTA-TCCAGAGAGCATGAAGAAGC 600
Qy      857  AGCT--GGGCGAGAGGCTTCAGGCTTCAC-TCCGCGGACTTT--GCCATCTCAATGC 911
Db      601  AACTTGGGGGCGAGAGGCTTCAACCTTACTTCCGCGGACTTTCGCACTCTTAAATGC 660
Qy      912  CGGAGAGAC 921
Db      661  CGGAGAGAC 670

```

## RESULT 9

```

US-09-417-507-9340
; Sequence 9340, Application US/09417507
; GENERAL INFORMATION:
; APPLICANT: KEITH G. WEINSTOCK ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS
; TITLE OF INVENTION: FUNGICIDUS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PAT99-10
; CURRENT APPLICATION NUMBER: US/09/417,507
; CURRENT FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 44312
; SEQ ID NO 9340
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: A.fumigatus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1562)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-417-507-9340

```

```

Query Match      23.7%; Score 345.4; DB 20; Length 1572;
Best Local Similarity 55.7%; Pred. No. 1.4e-68;
Matches 721; Conservative 2; Mismatches 563; Indels 9; Gaps 3;

```

```

Qy      27  CGACTTGTATGAGGCTTTCGCAACGCGCGCTTACAGATCGAAGGCGCCGTCAAGAAGG 86
Db      66  CGACTTGTATGAGGCTTTCGCAACGCGCGCGCTTACAGATCGAAGGCGCCGTCAAGAAGG 125
Qy      87  TGGCCGCGCCGTCATCTGGGACAGTACTGCCACTGAGGCATTCGCGCAACCAACG 146
Db      126  CGGTAAAGGCCAATCAATCTGGGATCTTTTGCACACACCTGGGAAAGTGAAGACGG 185
Qy      147  CGGCAAGGCGATGAGGCTTGCATCACTACACCGCTACGATGAGACTTTGATCTCTT 206
Db      186  AAGCATGAGGACAGCGCGGTACGCTGTATGACTCTGACAAAGAAATGTGGGCTCAT 245
Qy      207  GACCAAGTACGCGCAAGGCGCTACCGCTTCTCTGTGTGTGCGGATCATTTCCCT 266
Db      246  GAAATCGATATGAGGCTTCAATGATACCGCTTCTCTGTGTGTGCGGATCATTTCCCT 305
Qy      267  CGGCGGAGGCTGATCCGCTCAACGAGAGGAAATTTGATTAAGCAACGATGATGA 326
Db      306  TGGAGGGTGCACGACCTTCGCAAGAAAGGAGATTAATTAATTAATTTGATTTGA 365
Qy      327  CGCCTGTTGAGGCGGGGTATCAAGCTTGGGTGACTTTGTATCAACTGGGATCTGCCTCA 386

```



```

Db      366 CGAGCTCTCCGTAATGATCATCACCGTTTGTCTACTCTCTCCAGTGGACACACCGCA 425
Qy      387 GGGGCTTACAGATCGCTATGAGAGCTGCTCAACGTGGAAGAGTCCAGCTGACCTTGA 446
Db      426 ATGGCTGGAGATCGCTACGCGGCATGCTGAACAGAGAAAGTTGTGCGGACTTGT 485
Qy      447 GCGGTATGCGAGGTTGCTTTTGAACGTTTGGGGAACGATCCAGATCGATCAACCAT 506
Db      486 CAATTACGCTCGCTTGTGCTTTCAGAGACTCGGCGATCGGCTCAAGATTTGATCAGTT 545
Qy      507 CAACGAMCCCTGATTCAGGCACTATGATATGCAACCGGACGACACCGCCCGGAG 566
Db      546 CAACGAACTGTGTCTATACGCTGCTGATATGACCGGCGCTCCACGACCGGCGC 605
Qy      567 GAGCAGATTAAACAAGATCTCACCGAGGCAACCTGCCATGAGCCGTGCTGCTGG 626
Db      606 GTCTTCAATCCGGGACCGCAACGAGAGGCGACTGCTCAACGAGCATTCATCTTGC 665
Qy      627 AAAGCCCATGATCATGAGCCATGCGCGCGCTGCGCTCAACGAGGACCTTTCGCC 686
Db      666 GCACAGGAGCTGTGCTCCACGCGGACGCTGTCCCTCTTCAAGCAAGATTCAGCC 725
Qy      687 CTGCAAAAAGGCGCAGATCGGCACTTCGCTCAACGCGCACTATGAGCCCTGGGACAG 746
Db      726 CCAATCAGCAAGGACCATCGGCATCAAGCTTCAAGGGAACTGGTCCGACCATGGATGA 785
Qy      747 CAATGAGCTTCGGGACAAAGAGCTGCTGAGCGCAAGATGGAATTTCACTTGGCTGTT 806
Db      786 AGCGGACCCCTGATGATMACCGCGCTGAGCGCGCGCGGAGTTGCAAGATTTGCTGGTT 845
Qy      807 TGCATATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAAGAAGAGCTGGGCGA 866
Db      846 CCGGAGACCCCTCTTCAAGAGCGGCTGATTCGCGCTCGATGAGAGCCAGCTGGTGA 905
Qy      867 GAGGCTTTCAGCCCTCACTCCCGGGAATTTCCTCATCTCAATGCGGAGAGACCGACTT 926
Db      906 CCGTGTGCGCAAGTTTCAAGCCCGGAGAGT---CAAGCTGTGCTTGGAGCTCCGAGTT 962
Qy      927 CTACGCGATGAATTTACTACATCTCCAGTTTCGCGCGCACCTAGACGATCCCTCCCGA 986
Db      963 CTATGGCATMACTGTAACAGACTTCTTCTGTCMACACAGACACGCGCGGAGCAT 1022
Qy      987 GAGGACTATCTGGGCGCATTCATGAGCAACGAGAAATAGGACGCGACCCCGTTGG 1046
Db      1023 TAACGACCAAGAGGGAATGATGATTTCAATGATTTCAACAACATGATCTTCGAGG 1082
Qy      1047 CGAGGAGCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
Db      1083 CGAAGAGAGGACACGAGAGTGGCTGGAGCTGCCCTCGGGGCTTCCGAAAGCTTGA 1142
Qy      1107 CCGGGGTACGCGCTGTACGCGCAAGCCCATCTACATCAACGAGAACGATGCTGCTGCC 1166
Db      1143 TTGGATCTGTCTCGGTATCAGATGCGCATCTATGTACGAGAGAT---GCAAGAGC 1199
Qy      1167 TGAAGGAGAAATGAGCTGCGAGAGGCGCATGAGCCCTTCGCAATCGGTAAT 1226
Db      1200 CAAGGAGGAGACAGCCCTTCAACCAATGTTCTCAAGACCAATTTGCAATGAGTCTT 1259
Qy      1227 TGACTCGCATTTGACT---GCAATTCGAAGGCAATTAACGAGACGCGCTGCTGCA 1283
Db      1260 TGAAGGATATGTTGGGTGGGCACTGGCTGAGGCTCAAGAGGAGCGGATGATATCCG 1319
Qy      1284 GGGGTACTTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1318
Db      1320 CTCCTATTTTGGCTGCAATTCACCGATACTGGG 1354

```

RESULT 10  
 US-09-417-507-19268  
 ; Sequence 19268, Application US/09417507  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KEITH G. WEINSTOCK ET AL.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ASPERGILLUS  
 ; TITLE OF INVENTION: FUNGICIDUS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: PAT99-10  
 ; CURRENT APPLICATION NUMBER: US/09/417,507  
 ; CURRENT FILING DATE: 1999-10-14  
 ; SEQ ID NO 19268  
 ; LENGTH: 741  
 ; TYPE: DNA  
 ; ORGANISM: A.fumigatus  
 US-09-417-507-19268

Query Match 23.0%; Score 334.4; DB 20; Length 741;  
 Best Local Similarity 67.3%; Pred. No. 4,2e-66;  
 Matches 470; Conservative 1; Mismatches 227; Indels 0; Gaps 0;

```

Qy      621 CGCTGAAAGGCCCAATCATGAGCATGCGCGCGCGCTGAGCGGCTTACAGCGGACTT 680
Db      6 CGCTGTGCGGCCCATCCCGTGTGAATACCCGCGCGCTGCTGTACAAAGAGATT 65
Qy      681 TCGCCCTCGCAAAAGGCGCAGATCGGCATCTGCTCAACGGGCACTATGAGCCCTG 740
Db      66 CCGGTCTGTCAACAGAGGCAAGTATGAGATCTGTTGAACGAGACTATATAGCTTGG 125
Qy      741 GGAACGCAATGAGCTTCGGGACAAAGAGAGCTGCTGAGCGACGATGGAATTTCAATTGG 800
Db      126 GAATGCGAAGAGAGAGCGGATACCGCTGTGAGAGCGTCCGATGGAATTTCAATTGG 185
Qy      801 CTGCTTTGCCAATCCCATCTTCTTGAAGAAGACTATCCAGAGATGAAGAAGCAAGCT 860
Db      186 CTGCTTCGCAATCTCTGCTTCTTCTGACAGAGACTACCCAGCATGATGAGAGCAACT 245
Qy      861 GGGCGAGGCTTTCAGCCCTCACTCCCGGCACTTTCGCATCTCAATGCGGAGAGAC 920
Db      246 TGAAGCTCGCTACCAAGATTTTCCCTCTGTGATTTGACCTCTTCGCGAGCCGAGAG 305
Qy      921 CGACTTCTACGCGATGAATTTACTACATCCAGTTTCGCGCGCACCTAGACGATCCCT 980
Db      306 CGATTTCTACGGGATGAACTATACATCCAGTTTGTGCTGCGCATCGGATCAACAGC 365
Qy      981 CCCCAGACGGAATATCTCGGCGCATTCATGAGCACAGAGAAATAGAGCGGAGCCC 1040
Db      366 ATCCGAACAGATTAATTTGCAATGATGATGATTAAGAGAAAGATGAAGGAGCTTC 425
Qy      1041 CGTTGGCGAGGAAGGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1100
Db      426 AGTTGGGAAACCAAGGTGATTCATTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485
Qy      1101 TCTGCGCGGGGTGATCGCTGTACGCGCAAGCCCATCTACATCAACGAGAACGATGCC 1160
Db      486 CTTACAGAGAGTTATTCGCTGTACGGAACCAATTTTTCATCAACGAGAACGATGTCC 545
Qy      1161 GTGCGCTGAGAGAGAAATGAGCTGCGAGAGGCGCTGACGAGACCCCTTCGCAATCG 1220
Db      546 ATGTCTCGAGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 605
Qy      1221 STACTTTGAATCGCATTTGATCTGATTTCCAGGCGCATTAACGAGCGGCTGCTGCT 1280
Db      606 GTATTTGAGATCATCTAGAGGCTGTGGCTGTCTGTCAACCAAGAGCGTCTGATAT 665
Qy      1281 CAAGGGTACTTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318
Db      666 CAGAGGCTACTTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703

```

RESULT 11  
 US-09-404-520-15403  
 ; Sequence 15403, Application US/09404520  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Ghodseli, Aziza  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: McIninch, James

```

; APPLICANT: Timberlake, William E.
; APPLICANT: Yu, Jaehyuk
; TITLE OF INVENTION: Emericella nidulans Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15498)A
; CURRENT APPLICATION NUMBER: US/09/404,520
; PRIORITY FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 44345
; SEQ ID NO 15403
; LENGTH: 2918
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-09-404-520-15403

```

Query Match 21.9%; Score 318.6; DB 20; Length 2918;

Best Local Similarity 53.8%; Pred. No. 2,1e-62;

Matches 701; Conservative 1; Mismatches 595; Indels 6; Gaps 2;

```

QY 16 GCTCTGCCCAAGCATTTGATGGGCTTCGCAACGGCCGCTACAGATCGAAGCGCC 75
DB 1147 GCTTGGCCCAATGACTTCTTCATGATATGCAACGGCGCGCGAGTTGAAGAACG 1206
QY 76 GTCAAGAAGGAGCGCGGCGCCGCTCATCTGGGCAAGTACTGCACTGAGCATCG 135
DB 1207 TGAACAAGATGGCAAGGCTCTTGATCTGGGACACATTCGGCATATCTCAAGCANA 1266
QY 136 CGACCAACGGCGCGCGGATGCGCTTGCACTACACCGCTACGATGAGGAC 195
DB 1267 GTCAAGAAGCATATGCAATGCGGATGACGCTGAGGTTCTACACTTTACCGGAAT 1326
QY 196 TTTGATCTCTTGACCAATGAGCGCAAGGCTTACGCTTCTCTGTCGTGTCGG 255
DB 1327 GTTGGCCGATGAATGCTATGATGTGTCAACGCTTATCATTCGCTTCTGTCGGCG 1386
QY 256 ATCATTCCTCCGCGCGGAGGCTGGATCCGTCACAGAGGGAATGAATTTTACAC 315
DB 1387 ATATCCCACTCGAGGCGCGGACGACCGGTCATATGCAAGGATCAAGTACTACAG 1446
QY 316 AAATGATGAAGCGCTTGTAGGCGGGATCAACGCTTGGTGTACTTTTACACTG 375
DB 1447 GACCTGGTAGAGAGCTACTCATATATGATGATAGGATTCAGCATTCCTTTCACTG 1506
QY 376 GATCTGCTCAGGCGCTTCAAGATGCTATGAGAGCTGCTGCAACGTGGAAGAGTCCAG 435
DB 1507 GATGTTCCGAGGCTTGTGAGAGATCGGATCGGGGCAATGCTGAACCAAGAAAGTTCA 1566
QY 436 CTGACTTTGACGGGTATGAGGTTGCTTTGAAGCTTTTGGGAGCCGATCCAGAC 495
DB 1567 CTTGACTTTGCTGCGCTAGCGCGCTGTCTGTTTCAACGCTTACGAGCCAAAGGTCGTCAC 1626
QY 496 TGGATCAACATCAAGACCTGTGATTCAGGCACTATATGATATGCAACCGGACAC 555
DB 1627 TGGATTCATTCATATGAACCTGTGTGTATCTGCTGGAGGTTATGCGGAGGTTCAT 1686
QY 556 GCCCGGCGGAGAGAGATTAACAAGATCTCAACGAGGCAACATCTGCCATCTAGCCG 615
DB 1687 GCTCTGAGCGGTATGTTTGTGGAACCTCAATGAAGAGGTGATTCCTTCACAGAACCG 1746
QY 616 TGGCTGCTGGAAGAGCCAGATCATGAGCCATCCGCGCGCTGAGCGATTAAGCAGG 675
DB 1747 TTTATTCGTGCGGCAACAAACTGTGACCCACGCGCATGTTTCCAAAGCTTTACCGAG 1806
QY 676 GACTTTGCGCCCTGCAAAAGGCGCAAGTGGCATCTGCTCAACGCGCATCTATGAG 735
DB 1807 GTGTTCCAACCGGAGCAAGAAAGCAATTTGGCATCACTCCATGCAACTGTGCGAA 1866
QY 736 CCCGCGGAGAGCATATGAGCTCGGGAGCAAGAGGCTGTGAGCAGAGATGAATTTTAC 795
DB 1867 CTTTGGAGTGAAGAGCATCGCGGAGCACAGAGACACAGAGCGGCAAGGAAATTTGAG 1926
QY 796 ATTGCTGTGTTTGGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGATGAAG 855
DB 1927 ATGCTTGTGTTCCGAGACCTTTATACAGAGAGGCACTACCAAGCTCTCAATGCGGCT 1986

```

```

QY 856 CAGCTGGGAGAGAGGCTTCCAGCCCTCAGTCCCGCGGACTTTGCAATCTCATGCCGA 915
DB 1987 CAATCGGAGACGCTTACCGCGCTTTTACTCCGAGAGT---CAAACTGGTATCAAGGA 2043
QY 916 GAGACCCACTTCTTACCGCATGAATTAATCAACATCCAGTTCGCGCCACCTAGACG 975
DB 2044 AGTTCAAAATTTTACGATATGAACCTGTACACGACCTTTTTCGAGAGCAAGATACG 2103
QY 976 CCGCTCCCGAGAGAGACTATCTCGGCGCATTCATGAGACCAAGAGAAATTAAGAGCG 1035
DB 2104 CTTCAAGATATCAATGACCAAAAGCAATGTATTTTACATACAAACAGCAAGGCG 2163
QY 1036 AGCCCGCTTGGAGAGAGAGCGGCTTCCGCTGCTGCTCTGCGCGCATGTTCCG 1095
DB 2164 GTCTCTGAGAGCGAAGAGTCCGATAGCGCTTGGCTGCGACCGGACCATCTGATGCGC 2223
QY 1096 AAGCATCTGCGCGGCTGTACGCGCTGTGACGGAAGCCATCTACATCAACGGAACGA 1155
DB 2224 AAGCTGCTCAATTTGATTTGAACCGATATCATGTGCTATATATGTCAAGAAATG-- 2281
QY 1156 TGCCCGTCCCTGAGAGAGAGAACTAGACGTGAGAGAGGCGTCAACGACCTTCCG 1215
DB 2282 -GACGACAGCAAAAGAGAGACAGCTCCACACAGAGTACTCATCGACTTTCCGT 2340
QY 1216 ATCCGTAATTTGATCTGCACTTGAATTTCCAAAGCCATTAACCCAGAGCGCTC 1275
DB 2341 ATGCGATTTCTGAAAGCTATCGTTGATGATTTGGCCCGTCTCTCAAGAGATGAGTG 2400
QY 1276 GTGTCMAAGGGGTAATTTGCGTGGGCGTGTGCTGATTAATTTGG 1318
DB 2401 GATATCGGCTTACTTGTGCTGAGCTTCAACGACCAACTGAGG 2443

```

# RESULT 12

US-10-369-493-37092

; Sequence 37092, Application US/10369493

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 37092

; LENGTH: 1470

; TYPE: DNA

; ORGANISM: Aspergillus nidulans

; US-10-369-493-37092

Query Match 21.8%; Score 317.8; DB 50; Length 1470;

Best Local Similarity 53.8%; Pred. No. 3e-62;

Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;

```

QY 20 TGGCCAAAGCATTTGATGGGCTTTCGAACGGCGGCTTACAGATGAAGCGCGCTCA 79
DB 2 TGGCAATGATCTTTCATGATATGACAGCGGCGGCGAGGTTGAAGAGCGGTGA 61
QY 80 AAGAGGTGGCGCGCGCGCTGCTCATTTGGGACACGTAATCTGACCTGAGGACATGCGCA 139
DB 62 ACAAGATGGCAAGGCTCTTTCATCTGGGACACATTCGGGATATCTCCAGCAAGTCA 121
QY 140 CCAAGCGGCGCAAGCGCATGTGCTTGGCATCACTACACCGCTTACATGAGGACTTTG 199
DB 122 AAGACATATGCAATGGGATGACGCTGTGAGGTTCTACGACTTTTACCGCAAGATTTG 181
QY 200 ATCTTTGACCAAGTACGGCGCAAGGCGCTTACCGCTTCTGTTGTGTGTCGGATCA 259

```

Db 182 CCTGATGAGTGTATGTGTCAGCCTTATGATCTCGATCTCTGGTCGGGCAATA 241  
 Qy 260 TTCCCTCCGGGCGAGCTGATCCCTCAACGAGAGGAAATGATTTACAGCAAC 319  
 Db 242 TCCACTCCGAGGCGCGAGACCCGGTCAATAGCAAGGAATCAACTACTACAGACC 301  
 Qy 320 TGAATGACGCTGTGAGGCGGGGTATCAGCGCTTGGGTGACTTTGTACACTGGGATC 379  
 Db 302 TGGTACAGAGTACTACTATATATGGATTAAGCCATTCTGTAACCTTTTCACTGGATG 361  
 Qy 380 TGGCTGAGGCGCTTCAAGATCGTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGCTG 439  
 Db 362 TTCGAGGCTTTGAGAGATCGGATAGGGGGCAATGTCGAACAGAGAGGTTCACTTCG 421  
 Qy 440 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAAAGCTTTTGGGAGCCAGAGTCCAAACTGGA 499  
 Db 422 ACTTTGTCGGTACGCGCGGTGTCTGTTTCAAGCGCTTGAAGGCCAAGAGTCCGCACTGA 481  
 Qy 500 TCACCATCAACGAMCCCTGGAATTCAGGCACTATGGAATGACCGGAGCAAGCCG 559  
 Db 482 TTACATTCAATGAACCTGTGTGTATCTCGTGGCAGGTTATGCGGCAAGGTTCATGCTC 541  
 Qy 560 CGGCGAGAGACAGCATTTAACAGCACTCCAGAGGCAACACTGCCACTGAGCCGTGAC 619  
 Db 542 CTGACGGTATGCTGTTTCCGAACTCAATGAGAGGGTGAATTCCTGACAGAACCTTTCA 601  
 Qy 620 TCGCTGAAAGGCGCCAGATATGAGCATGCCCGCCGCGTGGCCGTTCACAGAGGACT 679  
 Db 602 TCGTCCGCCACACAAACCTGTGATCCACGCGCATGTTCCTCAAGCTTTTACCGAGAGTGT 661  
 Qy 680 TTGCGCCCTGCAAAAGGGCCAGATCGGCATCTCGCTCAAGCGGACTTACTATGAGCCCT 729  
 Db 662 TCCAACTCGCAGAAAGAAACAAATTTGGCATTCACCTTCATGGAATGTTGGAACTTT 721  
 Qy 740 GGGACAGCAATGAGCCTCGGAGACAGAGGCTGTGAGCAGCGATGGAATTTCACTTG 799  
 Db 722 GGGATGAGACGATCCGCGGAGACAGAGACAGAGAACGGGCGAGGAATTTGAGATCG 781  
 Qy 800 GCTGTTTGGCAATCCCATCTTTCTTGAAGAAGACTTCCAGAGAGCATGAAGAGAC 859  
 Db 782 CTGTGTTCCCGACCCCTTATACAGACAGGCGATACCCAGGCTCAATCGGGCTTCAC 841  
 Qy 860 TGGGCGAGAGGCTTCCAGCCCTCACTCCGCGGACTTGGCATCTCAATCGCGGAGAGA 919  
 Db 842 TCGGGAGCGTCTTACCGCGTTTACCGGAGAGT---CAAACTGTGATCTAGAAAGTT 898  
 Qy 920 CCGACTTCTACGCGCATGATTTACTACATCCAGTTGCGCGGCACTTGAAGCGTCCG 979  
 Db 899 CAGAAATTTTACGATATGAACTCGTACAGACCTTTTTCGTGAGCACAAGGATACGCTC 958  
 Qy 980 TCCCGAGACGAGCTATCTTGGCGCATTCATGAGCACAAGAGATTAAGAGCGGAGCC 1039  
 Db 959 CAGATATCATATGACCAAAAGGCAATGTCTATTTCTACGATACAAACAGCAAGGCTCT 1018  
 Qy 1040 CGTTGGCGAGAGAGGCGCTGCGCTGCTGCGCTCCTGCGGAGATTTCCGGAAGC 1099  
 Db 1019 CTGAGGCGAGAGTCCGATACCGCTTGGCTGCGAGCGGACCCACTGATGGGCAAGC 1078  
 Qy 1100 ATCTCGCCCGGTGTATCGGCTGTATCGGCAAGCCCATCTACATCACCGAAGCGATGCC 1159  
 Db 1079 TGGTCAATTTGATTTGGAAACGATATCATGTGCTATATATGTCAGGAAATG---GCA 1135  
 Qy 1160 CGTGCGCTGGAGAGAGAACATAGCTGTGAGAGAGCGCGTCAAGACCCCTTCGCACTCC 1219  
 Db 1136 CGACAGCAAAAGAGAGACAGCTCCCAACAGAGGATATCATGACACTTTCGTATGTC 1195  
 Qy 1220 GATACCTTGAAGCTGAGCTGATTCATTCAGAGCCATTTACCCAGAGCGGCTGCTG 1279  
 Db 1196 GATTTCTTCAAGGCTACGTTGTGTGATTTGGCCGCTGCTTCAAGAGAGATGAGTGTATA 1255  
 Qy 1280 TCAAGGGGATCTTTCGTGGCGTGTGCTCGATTAACCTTGG 1318

Db 1256 TCCGATCTTACTTTGGCTGAGCCTTCACGACAACTGGG 1294  
 RESULT 13  
 US-60-360-039-37092  
 ; Sequence 37092, Application US/60360039  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Chen, Xianfeng  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)A  
 ; CURRENT APPLICATION NUMBER: US/60/360, 039  
 ; CURRENT FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 37092  
 ; LENGTH: 1470  
 ; TYPE: DNA  
 ; ORGANISM: *Aspergillus nidulans*  
 US-60-360-039-37092  
 Query Match 21.8%; Score 317.8; DB 93; Length 1470;  
 Best local Similarity 53.8%; Pred. No. 3e-62;  
 Matches 699; Conservative 1; Mismatches 593; Indels 6; Gaps 2;  
 Qy 20 TGGCCAAAGCATTTGAATGGGGCTTGGCAAGGCGGCTTACAGATCGAAGGCGCGTCA 79  
 Db 2 TGGCGAATGACTTCTTCATGATATGCCAGCGCGCGGAGGTTGAAGAGCGTGA 61  
 Qy 80 AAGAAGTGGCGCGCGCGCGCTTCATGCTGGAGACAGTATGCCACCTGGAGACCTCGGCA 139  
 Db 62 ACAAGATGGCAAGGGTCTTTCATCTGGACACATTCGGGCAATTCACAGCAAAAGTCA 121  
 Qy 140 CCAAGCGCGCAACGCGAGATGTGGCTTGCATCTACCAACCGCTACGATAGAGACTTTG 199  
 Db 122 AAGACATAGCAATGGGATGAGCTGTGAGTTCATACGATTTTACCGGAAAGATGTTG 181  
 Qy 200 ATCTCTTGAACAGTACGGGCAAGGCTTACCGCTTCTCTTGTGCTGTGCGGATCA 259  
 Db 182 CCTGATGAAGTGTGTGTGTGTCACGCTTATGATTTCTGCTATCTCGTGTCCGCAATA 241  
 Qy 260 TTCCCTCCGGGCGAGCTGATCCCGTCAACAGAGAGGAAATGATTTACAGCAAC 319  
 Db 242 TCCACTCCGAGGCGCGGAGACCCGGTCAATGAGAGAGGATCAATACAGAGACC 301  
 Qy 320 TGAATGACGCTGTGAGGCGGGGTATCAGCGCTTGGGTGACTTTGTACACTGGGATC 379  
 Db 302 TGGTACAGAGTACTACTATATGGATTAAGCCATTCTGTAACCTTTTCACTGGATG 361  
 Qy 380 TGGCTGAGGCGCTTCAAGATCGTATGAGAGCTGGCTCAACGTGAGAGAGTCCAGCTG 439  
 Db 362 TTCGAGGCTTTGAGAGATCGGATAGGGGGCAATGTCGAACAGAGAGGTTCACTTCG 421  
 Qy 440 ACTTTGAGCGGTATGCGAGGTGTGCTTTGAAAGCTTTTGGGAGCCAGAGTCCAAACTGGA 499  
 Db 422 ACTTTGTCGGTACGCGCGGTGTCTGTTTCAAGCGCTTGAAGGCCAAGAGTCCGCACTGA 481  
 Qy 500 TCACCATCAACGAMCCCTGGAATTCAGGCACTATGGAATGACCGGAGCAAGCCG 559  
 Db 482 TTACATTCAATGAACCTGTGTGTATCTCGTGGCAGGTTATGCGGCAAGGTTCATGCTC 541  
 Qy 560 CGGCGAGAGACAGCATTTAACAGCACTCCAGAGGCAACACTGCCACTGAGCCGTGAC 619  
 Db 542 CTGACGGTATGCTGTTTCCGAACTCAATGAGAGGGTGAATTCCTGACAGAACCTTTCA 601  
 Qy 620 TCGCTGAAAGGCGCCAGATATGAGCATGCCCGCCGCGTGGCCGTTCACAGAGGACT 679  
 Db 602 TCGTCCGCCACACAAACCTGTGATCCACGCGCATGTTCCTCAAGCTTTTACCGAGAGTGT 661  
 Qy 680 TTGCGCCCTGCAAAAGGGCCAGATCGGCATCTCGCTCAAGCGGACTATGAGCCCT 729

Db 662 TCACCCGCGAGAGAGAGCAATTTGGCATACCTTCATGGCAATCGATCGGAACCTT 721  
Qy 740 GGGACAGCAATAGAGCTCGGGACAAAGAGCGTGTGAGCGACGAGTGAATTTTCACATTTG 799  
Db 722 GGGATGAAGAGATCCGCGGAGCCAGAGAGCAGAAACGGGATTCGAGATTCG 781  
Qy 800 GCTGTGTTCATCCATCTCTTTGGAAGAGACTATCCAGAGAGATGAAGAGAGC 859  
Db 782 CTGTGTTCGCGAGCCCTTTATCAAGACAGCGCATCCAGCTCAATGTGGGCTCAAC 841  
Qy 860 TGGGCGAGAGCTTCAGACCTCACTCCGCGGACTTTGCAATCCTCAATGCCAGAGAGA 919  
Db 842 TCGGGAGCGCTTACCGCGTTTCACTCCAGAGAGT---CAAACTGTACTAGAAATTT 898  
Qy 920 CCGACTTCTACGCGATGAAATTAACAATCCAGTTGCGCGCGCCACTTACAGCGTCCG 979  
Db 899 CAGAAATTTACGGTATGAATCTGTACAGACCTTTTTCGTGCAACAGAGATACGCTTC 958  
Qy 980 TCCCGGAGAGGACTATCTCGGCGCCATCCATGAGCACAGGAGAAATTAAGACGGCAGCC 1039  
Db 959 CAGATATCAATGACACAAAGGCAATGTCAATTTTCAAGATACAAACGCAAGGCGTCT 1018  
Qy 1040 CCGTTGGCGAGAGAGCGGCGCTCGCTGCGCTCGCTCGCGAGATGTTCCGAGAGC 1099  
Db 1019 CTCGAGCGAGAGATCCGATACGCTTGTGCGCGAGCGGACCCACTGATGGGCGAAGC 1078  
Qy 1100 ATCTCGCCCGGGTGTACGCGCTGTACGCGCAAGCCCATCTACATCAACGAGACGATGCC 1159  
Db 1079 TGCTCAATTTGATTTGGAAACGATATCATGTGCTATATATGTACGGAATAATG---GCA 1135  
Qy 1160 CGTCCCTGAGAGAGAGAACTAGACGTGCGAGAGGCGGTCAAGAGACCCCTTCGCAATCC 1219  
Db 1136 CGACAGCAAAAGAGAGACAGCTCCACACCAAGAGATCTATGACACTTTTCGGTATGC 1195  
Qy 1220 GGTACTTTGACTCCGACTGTGACTCGATTTTCCAAAGCCATTAACCAAGACGCGTGTGCG 1279  
Db 1196 GATCTTCGAAAGGCTAGGTTGTGGATTTGGCCCGGTGCTGCAAGAGAGATGAGATGATA 1255  
Qy 1280 TCAAGGGGATCTTTGCGTGGCGCTTCTCTCAATTAATTGG 1318  
Db 1256 TCCGCTTACTTTGCTGCTGACCTTCAACCGAACCTGGG 1294

RESULT 14  
US-60-138-103-7477  
; Sequence 7477, Application US/60138103  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Timberlake, William E.  
; TITLE OF INVENTION: ASPERGILLUS NIDULANS GENOME SEQUENCE AND USES THEREOF  
; FILE REFERENCE: 38-101154851D  
; CURRENT APPLICATION NUMBER: US/60138,103  
; NUMBER OF SEQ ID NOS: 28006  
; SEQ ID NO 7477  
; LENGTH: 2914  
; TYPE: DNA  
; ORGANISM: Aspergillus nidulans  
US-60-138-103-7477

Query Match 21.1%; Score 307.6; DB 69; Length 2914;  
Best Local Similarity 53.8%; Pred. No. 7,1e-60;  
Matches 701; Conservative 1; Mismatches 595; Indels 7; Gaps 3;

Qy 16 GCTGTGCCCAAGCATTTGAATGGGCTTTCGCAACGCGCGCTTACAGATGCAAGCGGCC 75  
Db 1147 GCTTGGCCCAATGACTTCTTCATGATATGCAAGCGCGCGCGCAGAGTTGAAGAGCG 1206  
Qy 76 GTCAAGAGAGTGGCGCGCGCGCTTCATCTGAGGACAGTACTGCACTGAGGCAATCG 135  
Db 1207 TGGAAACAAGATGCAAGGCTCTTCGATCTGGGACACATTCGGGCAATCTCCAGGCAAA 1266

Qy 136 CGACCAAGCGCGCCCAACGCGATGTGGCTTGGCATCACTACCAACCGCTACGATGAGAGC 195  
Db 1267 GTCAAGAGCATATGCAATGGAGATGAGCGCTGTGAGGTTCTACGACTTTTACCGCAAGAT 1326  
Qy 196 TTTGATCTCTTGAACCAAGTACGCGCAAGAGCTTACCGCTTCTCTTGTGTGTGCGG 255  
Db 1327 GTTGCCCTGATGAAGTCGTATGTGTCAACGCTTATCTGATTCCTGATTCCTGTGCGCG 1386  
Qy 256 ATCATTTCCCTCGGCGCGAGCGTGTGATCCCGTCAAGAGAGGAGAAATGATTTTACAGC 315  
Db 1387 ATATCCCATCTGGAAGCGCGCGAGACCGGTCAAATGAGAGAGAGATCAATGATCTACAG 1446  
Qy 316 AAATGATTTGACCGCTGTGAGCGGAGTATCAAGCGCTTGGGTGAATTTTGAACAATGG 375  
Db 1447 GACTGTGTAAGAGGATCACTCAATATATGGAATTAAGCAATTCGTAACCTCTTCACTGG 1506  
Qy 376 GATTCGCTTACGCGCTTACAGATTCGCTATGAGAGCTGCTCAACGTGGAAGAGTCCAG 435  
Db 1507 GATGTTCCGAGGCTTTGAGAGATCGGTACGCGGGGATGCTGAACCAAGAGAGGTTCAAT 1566  
Qy 436 CTGACCTTGAAGGGTATGCGAGGTTGTGCTTTGAAGCTTTTGGGACCGAGTCCAGAAC 495  
Db 1567 CTGACTTTGTCCGCTACGCGCGGTGTCTTTGCAAGCCTTAAGGCCAAAGTCCGTCAC 1626  
Qy 496 TGAATCACATCAACAGMCCCTGG-ATTACAGGCATCTATGATATATGCAACGCGAGCAA 554  
Db 1627 TGGATTAATTAATGAAATCATGTGTGTATCTGCTGCGAGGTTATGCGCAGAGTTTCA 1686  
Qy 555 CGCCCCGGGAGAGAGACATTTAAACAAGCACTCAACGAGGAGCAACATGCGCATGAGCC 614  
Db 1687 TGCTCTGACGGGTATCGTGTTCGCGAACTCAATGAAGAGGGATTCCTGACAGAAAC 1746  
Qy 615 GTGGCTGCTGGAAGAGCCCAAGATCATGAGCAATGCCCGCGCGGTGCGCTTACAGCAG 674  
Db 1747 GTTCATGTGTGCGCACACAAATGTGTGACCCAGCGCATGTTTCCAGCTTTATCCAGAGA 1806  
Qy 675 GAACTTTGCGCCCTCGCAAAAGGCGCAGATCGGATCTGCTCAACGCGCACTATATGA 734  
Db 1807 GGTGTTCACACCGAGAGAAAGAAACATTTGGCATCACTCCATGAGCAATCTGTGCGGA 1866  
Qy 735 GCGCTGGGACAGCATATAGCTTGGGACAGAGAGCTGTGAGAGCGAGTGAATTTCA 794  
Db 1867 ACCTTGGGATGAAGACATCTCGGAGCCAGAAAGACAGACCGGCGCCAGGGAAATTCGA 1926  
Qy 795 CATGCGCTGTTTGGCAATCCATCTCTTGAAGAGAACTATCCAGAGAGCATGAAGAA 854  
Db 1927 GATGCTTGTGTCCCGAACCCCTTATCAAGACAGGAGACTACCAACCTCAATGCGGGC 1986  
Qy 855 GCACTGTGGGAGAGAGGCTTCAAGCTTCACTCCCGGAGCTTGGCATCTCAATGCCCG 914  
Db 1987 TCAACTCGGGGACGCTTACCGCGTTCACCTCCGAGAGAT---CAAACTGTGATTAAG 2043  
Qy 915 AGAAGCCGATCTTCAAGGCAATGAAATTTACTACATCTCCAGTGTGCGGCGCACTTGAAG 974  
Db 2044 AAGTTCAAGAAATTTTACGATGAACCTGTACACACACTTTTGTGTGAGCAACAAGATAC 2103  
Qy 975 TCCGCTCCCGAGAGCGGACTATCTCGGCGCATTCATGAGACAGGAGATTAAGACG 1034  
Db 2104 GCTTCAGATATCAATATGACCAAGAGCAATGATATGTTTCAAGATCAAAACAGCAAGG 2163  
Qy 1035 CAGCGCGTGTGGAGAGAGAGCGGCTCGGCTGTGCTGCTCTCTCCCGACATGTTCCG 1094  
Db 2164 CGTCTTCGAGGCGAAGAGTCGATACGCTTGTGCTCGGACGCGACCCACTGTAGTGGC 2223  
Qy 1095 GAACATCTCGCGCGGTGTATAGGCTGTATCGGCAAGCCCATCTACATCAACGGAACG 1154  
Db 2224 CAAGCTGTCAATTTGAATTTGAACCAATATCATGTGCTTATATATGTCAAGAAAAATG- 2282  
Qy 1155 ATGCGCGTCCCTGAGAGAGAAATATGACGTGCGAGAGCGCTTCAACGACCCCTTCCG 1214  
Db 2283 --GACCAACAGCAAAAGAGAGACAGCTCCACACAGAGATTACTATGACACTTTCG 2340  
Qy 1215 CATCCGTAATTTGATCTGCACTTGAAGTCAATTTCCAGGCGCATTAACCAAGAGCGGCT 1274

Db 2341 TATGCGATTCCTCGAAGCTACGTGATGGCCCGTGTCTCAAGAGATGAGT 2400  
 Qy 1275 CGTCGTAAGGGGACTTTCGCGGGGCTGCTCGAATCTTG 1318  
 Db 2401 GGATATCCGGTCTTACTTGTGCTGACCTTACCGACACTGG 2444

## RESULT 15

US-09-404-520-19929  
 ; Sequence 19929; Application US/09404520  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Ghodessi, Azita  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: McIninch, James  
 ; APPLICANT: Timberlake, William E.  
 ; TITLE OF INVENTION: Emerice11a nidulans Genome Sequence and Uses Thereof  
 ; FILE REFERENCE: 38-10(15498)A  
 ; CURRENT APPLICATION NUMBER: US/09/404,520  
 ; NUMBER OF SEQ ID NOS: 44345  
 ; SEQ ID NO 19929  
 ; LENGTH: 1434  
 ; TYPE: DNA  
 ; ORGANISM: Aspergillus nidulans  
 US-09-404-520-19929

Query Match 20.2%; Score 294.6; DB 20; Length 1434;  
 Best Local Similarity 53.7%; Pred. No. 6,2e-57;  
 Matches 676; Conservative 1; Mismatches 575; Indels 7; Gaps 3;

Qy 61 CAGATCGAAGCGCCCTGCAAGAAAGTGGCGCGCCCTGCTCATCTGGGACAGTATGC 120  
 Db 3 CAGGTGAAAGAGCGTGGAAAGAGATGGCAAGGCTCTTCGATCTGGGACACTTCGGG 62  
 Qy 121 CACCTGAGCGCATCGCCGCAAGAGCGGCGCAAGCGGATGGCTTCGATCACTACAC 180  
 Db 63 CATATCTCGAAGCAAGTCAAGACATATGCAATGCGATACGCTTGAGAGTTCTAGAC 122  
 Qy 181 CGGTACGATGAGACTTTGATCTTGTACCAAGTACGCGCAAGGCTTCTCC 240  
 Db 123 TTTATCGCGAAGATGTGCTGATGATGATGATGATGATGATGATGATGATGATG 182  
 Qy 241 TTTGCTGCTGCGGATCATTTCCCTGCGGCGAGGCTGATCCCTGCAAGAGAGGA 300  
 Db 183 CTATCTGCTGCGGCAATTCCTGAGGCGCCGACGACCCGCTCAATGAGCAAGG 242  
 Qy 301 ATGAGTTTACAGCAAGATTTGAGCGCCTGTTGAGGCGGGTATCAGCCCTTGAGTG 360  
 Db 243 ATCAAGTACTACAGAGCTGTGATGAGAGCTACTCAATATGAGATTTAGCCATTGTA 302  
 Qy 361 ACTTTGACCATGAGATCTGCTCAAGCGCTTCAAGATGCTATGAGGCTGCTCAC 420  
 Db 303 ACCCTCTTACATGGAGTGTTCGCAAGGCTTTGAGAGATGAGTACGGGGCATGTAAC 362  
 Qy 421 GTGGAAGAGTCCAGCTGAGCTTTGAGCGGATGCGAGTTGCTTGAAGTTTGGG 480  
 Db 363 CAGGAGAGGTTCACTCTGACTTGTGCTGACGCGCTGCTGTTGAAAGCTAGGG 422  
 Qy 481 GACCGAGTCCAGAACTGATCACCATCAAGACCCCTG--ATTGAGGCACTTATGATA 539  
 Db 423 CCAAGGTCCTGATCTGAGATTTACATTTCAATGAAACATGTGTGACTGCTGGAGGTTA 482  
 Qy 540 TGGCACCAGAGCAAGCCCGGAGAGAGAGAGAGATTTAGCAAGCACTTCAAGAGGCAA 539  
 Db 483 TGGGAGAGGTTTATGATCTCTGAGAGGTATGCTTTTGGAGTCAATGAGAGGTTGA 542  
 Qy 600 CATGTCATGAGCGCTGCTGCTGAGAAAGGCCAGATCATGAGCATGCGCGCGCT 659  
 Db 543 TTCTCGACGAAGACCTTTCATCTGTGCGCACACAAACTGATGACCAAGGCAATGTTTC 602

Qy 660 GGCCTCTACAGAGGACTTTTCGCCCTTCGCAAAAGGCGCAGATCGCATCTCGTCAA 719  
 Db 603 CAAGCTTTACGAGAGGTGTTTCCACCGACAGAGAAAGAAATTTGGATCAACCTCCA 662  
 Qy 720 CGGAGTACTATGAGCCCTGCGAGCAAGATGAGCTCGGAGACAGAGGCTGTAAGG 719  
 Db 663 TGGCAACTGCTCGAAGCTTTGGATGAGAGCATTCGCGGAGCAAGAGCAGAGCCG 722  
 Qy 780 ACGGATGAAATTTACATTTGGCTGTTTGGCAATCCATCTTCTTGAAGAGACTATCC 839  
 Db 723 GGCAGGGAATTTGAGATGCTTGTGTTCCGAGCCCTTATACAGACAGGCACTATCC 782  
 Qy 840 AGAGAGATGAAAGAGAGCTGCGAGAGAGGCTTCCAGCCCTCACTCCCGAGACTTTCG 839  
 Db 783 AGCTCAATGCGGGCTCACTGCGGAGCGGTCAACCGCTTCACTCCGAGAGT---C 839  
 Qy 900 CATCTCAATGCGGAGAGAGCCGATCTTACGCGATGATTTACTACATCCAGTTGCG 959  
 Db 840 AAAACTGTACTAGGAAGTTCAAGAAATTTACGGTATGAACTCGTACACGACTTTTTCGT 839  
 Qy 960 GCGCCACTAGAGGTCCTCCGCGAGACGACTATCTGCGGCGCATCATGAGACCA 1019  
 Db 900 GCAGACAGAGATACGCTTCAGATATCAATGACCAAGAGCATGTCTATGTTACGA 959  
 Qy 1020 GGAAGATTAAGAGAGGAGCCCGTGGAGAGAGAGAGAGGCTGCGTGGCTGCTCTG 1079  
 Db 960 TACAAACAGCAAGAGGCTCTCTGAGGCGAAGAGTCCGATACGCTTGGTGGAGCGG 1019  
 Qy 1080 CCGGACATGTTCCGAAAGCATTCGCGCGGCTGTAAGGCTTGTACGAGCCCATCTA 1139  
 Db 1020 ACCCACTGATGCGCAAGCTGCTCAATGATTTGAAACCGATATCATGTGCTATATA 1079  
 Qy 1140 CATACCGAAGAGAGTCCCGGCTGAGAGAGAGAAACATACGTCGAGAGAGCGCT 1199  
 Db 1080 TGTACGAAATG---GACGACAGCAAAAGAGAGACGCTCCACACCAAGAGTACT 1136  
 Qy 1200 CAAGACCCCTTCGATCGSTACTTGTGACTGCACTTGGACTCGATTTTCAAGGCAAT 1259  
 Db 1137 CATGACACTTTTCCGATGAGATTTTCAAGGCTACGTTGAGATTTGSCCGGTGCTGT 1196  
 Qy 1260 TACCAAGAGAGCGCTGCTCAAGAGGATCTTTGCGTGGCGTTGCTCGATTAATTGG 1318  
 Db 1197 CAAGAGAGATGAGTGAATATCCGATCTTAATTTGCTGGAACCTTACCGACCAACTGGG 1255

Search completed: March 29, 2004, 02:51:02  
 Job time : 5017.04 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 00:15:12 ; Search time 124.98 Seconds

(without alignments)  
6942.044 Million cell updates/sec

Title: US-10-026-140-3

Sequence: 1 atgcccagctcgtcagctctc.....gggttaagtcgagcgtacaa 1455

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1264548 seqs, 298149298 residues

Total number of hits satisfying chosen parameters: 2529096

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents NA, New:\*

- 1: /cgn2\_6/prodata/2/pna/PCT\_NEW\_COMB.seq:\*
- 2: /cgn2\_6/prodata/2/pna/US06\_NEW\_COMB.seq:\*
- 3: /cgn2\_6/prodata/2/pna/US07\_NEW\_COMB.seq:\*
- 4: /cgn2\_6/prodata/2/pna/US09\_NEW\_COMB.seq:\*
- 5: /cgn2\_6/prodata/2/pna/US10\_NEW\_COMB.seq:\*
- 6: /cgn2\_6/prodata/2/pna/US60\_NEW\_COMB.seq:\*
- 7: /cgn2\_6/prodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	178.2	12.2	2016	6 US-10-152-372-253	Sequence 253, App
2	162.8	11.2	1374	1 PCT-US04-02242-95	Sequence 95, Appl
3	159.8	11.0	805	6 US-10-767-701-7585	Sequence 7585, Ap
4	151.6	10.4	6274	6 US-10-767-471-414	Sequence 414, App
5	147.8	10.2	767	6 US-10-767-701-6851	Sequence 6851, Ap
6	139	9.6	1389	1 PCT-US04-02242-93	Sequence 93, Appl
7	131	9.0	2042	6 US-10-767-701-14696	Sequence 14696, A
8	130.8	8.9	1338	6 US-10-093-037A-2	Sequence 2, Appl1
9	130.2	8.9	1338	6 US-10-767-701-6014	Sequence 6014, Ap
10	129.6	8.9	61330	6 US-10-767-471-10690	Sequence 10690, A
11	129.4	8.6	689	6 US-10-767-701-8210	Sequence 8210, Ap
12	119.4	8.2	649	6 US-10-767-795-4264	Sequence 4264, Ap
13	115.4	7.9	716	6 US-10-767-701-12773	Sequence 12773, A
14	114.8	7.9	637	6 US-10-767-701-26695	Sequence 26695, A
15	109	7.5	558	6 US-10-767-795-4474	Sequence 4474, Ap
16	108.8	7.5	2000	6 US-10-767-701-14973	Sequence 14973, A
17	107.4	7.4	348	6 US-10-767-701-10892	Sequence 10892, A
18	103.8	7.1	650	6 US-10-767-701-2039	Sequence 2039, Ap
19	102.4	7.0	1956	6 US-10-767-701-14945	Sequence 14945, A
20	101.4	7.0	1437	6 US-10-417-884A-1268	Sequence 1268, Ap
21	101.2	7.0	1312	6 US-10-767-701-5404	Sequence 5404, Ap
22	97.7	6.7	1310	6 US-10-767-701-13239	Sequence 13239, A
23	95.6	6.6	381	6 US-10-767-795-2194	Sequence 2194, Ap
24	89.2	6.1	487	6 US-10-767-701-30963	Sequence 30963, A
25	88	6.0	820	6 US-10-767-701-3304	Sequence 3304, Ap
26	73	5.0	552	6 US-10-767-701-26684	Sequence 26684, A

27	72	4.9	522	6 US-10-767-701-702	Sequence 702, App
28	67	4.6	127602	6 US-10-765-790-65	Sequence 65, Appl
29	61	4.2	476	6 US-10-767-701-23562	Sequence 23562, A
30	60.8	4.2	201	6 US-10-767-471-7895	Sequence 7895, Ap
31	60.8	4.2	201	6 US-10-767-471-29817	Sequence 29817, A
32	60.8	4.2	470	6 US-10-767-701-22496	Sequence 22496, A
33	60.8	4.2	641	6 US-10-767-701-29656	Sequence 29656, A
34	58.4	4.0	565	6 US-10-767-701-31359	Sequence 31359, Ap
35	57.2	3.9	613	6 US-10-767-701-31359	Sequence 31359, A
36	55.4	3.8	498	6 US-10-767-701-6735	Sequence 6735, Ap
37	54.2	3.7	1314	1 PCT-US04-02242-103	Sequence 103, App
38	53.2	3.7	708	6 US-10-767-701-1064	Sequence 1064, Ap
39	52.8	3.6	718	6 US-10-767-701-3709	Sequence 3709, Ap
40	51.4	3.5	495	6 US-10-767-701-22851	Sequence 22851, A
41	49.8	3.4	1239	6 US-10-767-701-5791	Sequence 5791, Ap
42	49.6	3.4	1512	6 US-10-417-884A-1359	Sequence 1359, Ap
43	49.4	3.4	800	6 US-10-767-701-8171	Sequence 8171, Ap
44	48	3.3	1263	1 PCT-US04-02242-89	Sequence 89, Appl
45	48	3.3	1365	6 US-10-093-037A-6	Sequence 6, Appl1

## ALIGNMENTS

RESULT 1  
US-10-152-372-253  
Sequence 253, Application US/10152372  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Luca  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Mei-Qiang  
APPLICANT: Geriltsen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P330R1C394  
CURRENT APPLICATION NUMBER: US/10/152,372  
CURRENT FILING DATE: 2002-05-21  
PRIOR APPLICATION NUMBER: 60/049911  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059586  
PRIOR FILING DATE: 1997-09-19  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 550  
SEQ ID NO 253

LENGTH: 2016  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-152-372-253

Query Match 12.2%; Score 178.2; DB 6; Length 2016;  
Best Local Similarity 54.5%; Pred. No. 2.3e-38;  
Matches 441; Conservative 1; Mismatches 334; Indels 33; Gaps 3;

```
QY 15 AGCTGCGCCCAAGCTTGAATGGGCTTCCCAACGGCCGCTTACGATGAAAGCGC 74
DB 237 AACCTTCCCTCTGGGCTTCTCTGGGCGCTGGGCGCTTCTGCTTACGACGAGGCGC 296
QY 75 CGTCAAGAAGGTGGCCCGCGCCGCTCCATCTGGGACAGTACTCCACCTTGGACCATC 134
DB 297 CTGGGACGAGACGGGAAAGGGCTTAGCATCTGGGACGCTTTCAACACAGTGGGAAAGG 356
QY 135 GCGCACCAAGG---CGCCACGCGGATGTGGCTTGCATCACTACCAACCGCTACGATGA 191
DB 357 GAAAGTGTCTGGGAATGAGACGCGCAGATGTAGCTGTACGCGCTACTCAAGGCTCAGGA 416
QY 192 GGACTTGTATCTTGAACCAAGTAGCGGCAAAAGGCTTACCGCTTCTCTGTGCTGTC 251
DB 417 GGAATCATCTTCTGAGGGAATCTGACGTCAACCACTACCGATTTCTCTGTCTTGGCC 476
QY 252 GCGGATCATTTCCCTCGGCGGCGGCTGGATCCGCTCAACGAGGAGGAAATTTGATTTTA 311
DB 477 CCGGCTCTGCGCCACAGGCGATCCGAGCGGAGGTGAAACAAAGGAAATCGAATTTCTA 536
QY 312 CAGCAAACTGATATGACGCGCTTGTAGGCGGGGATATCAACGCTTGGGTGACTTGTATCA 371
DB 537 CAGTATCTTATCGATGCGCTTCTTGAAGCAAACTCACTCCATCTGACCTTGGACCA 596
QY 372 CTGGGATCTGCTTCAAGGCGCTTCAAGATCGCTATGAGAGGCTTCAAGGCGGAAAGGT 431
DB 597 CTGGGATCTGCGACAGCTGCTTCCAGGTCAAAATACGATGGGTGGAGAAATGAGCANTGC 656
QY 432 CGAGTGAATTTGAGCGGATATGCGAGGTGTGCTTTGAAAGTTTGGGGAACGAGTCCA 491
DB 657 CAATGCTGAGACTAGACCTACGCAACCTGTGCTTGAAGCCTTGGGGAACCGTGTAA 713
QY 492 GAATGATTCACCAATCAACGAMCCTTGATTCAGGCCATCTATGATATGCGACCGGAG 551
DB 714 GCATGATGATCACTTCAATGATCTTCTGGGCAATGCGAAMAAAGCTATGAGACGGGCA 773
QY 552 CAAGCGCCCGGCGAGGACGACATTTAACAAGCACTCCACGAGGGAACACTGCACTGA 611
DB 774 CCAATGCGCGGGGCTGAAAGCTCCG-----GGACCGG 806
QY 612 GCCGTGCTCGTGAAGAGGCCAGATCATGAGCCGCGCGCGCTGCGCTCTACAG 671
DB 807 CCTGTACAAAGGACACACATCATTTAAGGCCCAAGAACTGCGACTTCTTAA 866
QY 672 CAGGGAATTTGGCCCTCGCAAAAGGCGCAGATGAGGATCTGCTCAACGCGCACTACTA 731
DB 867 CACCACTGCGGCGCAAGCAAGAGCTGTGTGGAAATTTCACTTGAATCTGTGCTGGG 926
QY 732 TGAGCCCTGGGACGACATGAGCTTGGGACAAAGAGGCTGCTGAGCGACGATGAAAT 791
DB 927 GGAACCTGTGACATTTGTAACCCCAAGGACCTTAGAGGCTGCGAGATATCTACGTT 986
QY 792 TCACATTTGCTGTGTTTGGCAATCCCATCT 820
DB 987 CTGTCTGGGCTGTGTTCACCAACCCCATTT 1015
```

RESULT 2  
PCT-US04-02242-95  
Sequence 95, Application PC/TUS0402242  
GENERAL INFORMATION:  
APPLICANT: Diversa Corporation  
APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine  
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN

TITLE OF INVENTION: THEM  
FILE REFERENCE: 56446-20118.40  
CURRENT APPLICATION NUMBER: PCT/US04/02242  
CURRENT FILING DATE: 2004-01-26  
PRIOR APPLICATION NUMBER: 60/442,794  
PRIOR FILING DATE: 2003-01-24  
NUMBER OF SEQ ID NOS: 224  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 95  
LENGTH: 1374  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
PCT-US04-02242-95

Query Match 11.2%; Score 162.8; DB 1; Length 1374;  
Best Local Similarity 60.0%; Pred. No. 2.8e-34;  
Matches 329; Conservative 0; Mismatches 207; Indels 12; Gaps 3;

```
QY 20 TGCCCAACGACTTTGAATGGGCTTCCCAACGGCCGCTTACGATGAAAGCGCGTCA 79
DB 38 TTCCCGCACCTTCACTGGGCGGTGGCCACAGCGCTTACGATGAAAGCGCGCGC 97
QY 80 AAGAAGGTGGCGGCGCGCTTCCATCTGGGACAGTATGCGCACTGAGGACCATTCGCA 139
DB 98 CCAATGGGCGCGCGCGCTTCCATCTGGGATCTTCAAGCAACAGGAGGAGGAGGAGG 157
QY 140 CCAAGGCGGCAACGCGGATGTGGCTTGGCATCTACCAACGCTACGATGAGACTTTG 199
DB 158 TCAGCGGCAACATGAGGAGGAGGCTGCGGACGACACTACCAACGCTATGCGGAGAG 217
QY 200 ATCTTGAACAGTATGAGGCGCAAGGCTTCAAGGCTTCTCTGTGTGGTGGGATCA 259
DB 218 AGCTGATGCGACCTTGGGCGTGAAGCGCTTACCGCTTTTCATGTCCTGTGCGGCTCC 277
QY 260 TTCCCTCGGCGGCGAGGCTGATCCGCTCAACGAGGAGGAAATTTGATTTACGAAAC 319
DB 278 AGCCCAAGGTTTC-----GGGCTTGAAGCAACAGGCTTTGATTTCTAGCCGCC 331
QY 320 TGAATGACGCTTGTGAGGCGGGGATATCAAGCTTGGGTGATTTTACACTGCGGATC 379
DB 332 TGCTCAAGCGCTTGGCGGCGCAAGGAGCTGACGCGCACTGACCTGTACACTGCGGACC 391
QY 380 TGCTCAGGCGGCTTCAAGATCGCTATGAGGCGGCTGCTCAACGATGAGAGTCACTGG 439
DB 392 TGCCGAAGCTTTCAGGA---CGAGGGCGGCTGGCTCAATGCGCCACTGTCTACC--- 445
QY 440 ACTTTAGCGGATATGAGAGGTGTGCTTGAACGTTTGGGACCGAGTCCAGAACTGGA 499
DB 446 ACTTGGCGGATATGCGCGCGAGGTGGCGCGGCTTGGCCACAGATGCGCAGATGG 505
QY 500 TCACATTCACAGMCCCTTGATTCAGGCGCATCTATGATATGCAACGCGCAGCAACGCC 559
DB 506 CCAAGCAATAGAGCGGTGTGATGCACTGCTGTGGCCACGAGCAGCGCGCATGTTCCGC 565
QY 560 CGGCGAGG 567
DB 566 CCGGCAATG 573
```

RESULT 3  
US-10-767-701-7585  
Sequence 7585, Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(53535)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29



NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 7585  
LENGTH: 805  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS93553\_1  
US-10-767-701-7585

Query Match 11.0%; Score 159.8; DB 6; Length 805;  
Best Local Similarity 54.3%; Pred. No. 1.5e-33;  
Matches 345; Conservative 0; Mismatches 287; Indels 3; Gaps 1;

20 TGCCCAACGACTTGAATGGGCTTCGCAACGCGCCGCTACAGATGGAAGCGCCCTCA 79  
174 TCCCAAGAGACTTGTCTTGGGACAGAGCTCGGGGCTATCAGTAGAGGCTCTACA 233  
80 AAGAAAGTGGCCGCGCCCTCCATCTGGGACAGTACTGCACTGCACTGGAGCCATCGCGCA 139  
234 ACGAAGAGGCAAGGGGCTACCATATGGGACAAATTCACTCATCTCCAGGTAAAGATCT 293  
140 CCAAGCGCCCAACGGGATGTGCTTGGCATCTACCAACGCTACGATGAGACTTTG 199  
294 TGAACAACGATACCGGACGATGACATGATGATCACTGATCAAGAGAGATGTC 353  
200 ATCTTGAACCAAGTACGCGGCAAGGCTACGCTTCTCTGCGTGGTGGGATCA 259  
354 AACTCTCAAGACATATAGAGCTTCGCGCTTCATCTGCTGACCAAGATCC 413  
260 TTCCCTCGGCGGACGCTGATCCGTCACAGAGGGAATTGATTTTACAGCAAC 319  
414 TGCACAAAGGTTCCCTGAGTGAAGGATCAACAAAGAGGAGTGGCTTCTATACAAAC 473  
320 TGAATGACGCTTGTAGGCGGGGATTCAGGCTTGGGATCTTTGATCAGCTGGATC 379  
474 TCAATCAACGAGGTCATAGCCAAAGGCTGAAGCATTTGTACCATTTCCATGAGACA 533  
380 TGCTCAGGCGCTTACGATCGCTAGAGGCTGGCTCAAGTGAAGAGGTCCAGCTGG 439  
534 CGCCCTCGGCGCTGAGAGCAAGATGAGAGATTCCTCAGGAGAACATCTCA---GG 590  
440 ACTTGAAGCGTATGCGAGGTTGCTTTGAAAGTTTGGGAGACGAGTCAGAACTGA 499  
591 ACTACGGAAGCTTGGCGAGGTTGCTTCAAGAGTTCCGGGACCGGCTCAAGGCTGGA 650  
500 TCACCATCAACGAGCCCTGATTCAGGCTATGATGATGCAACGGGACCAACCC 559  
651 CCAAGTTCACGAGCCGCTGACGTAATGTAACGAGGCTACGCGCTGCGCAAGTCGCGC 710  
560 CGGCGAGGACAGCATTAACAGACATCCACCGAGGCAACATGCACTGAGCGCTGGC 619  
711 CGGCGCGCTGCTGCTGATGATCAACAGAACTGCTTCCCGGCGACTCGGCGCGAGC 770  
620 TCGCTGAAAGGCCAGATGATGAGCCATGCGCCG 654  
771 CTTACACCGTGAAGGACACATATCTCTCGCCAC 805

RESULT 4  
US-10-767-471-414  
Sequence 414, Application US/10767471  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele et al.  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
FILE REFERENCE: CLO01505  
CURRENT APPLICATION NUMBER: US/10/767,471  
CURRENT FILING DATE: 2004-01-30  
NUMBER OF SEQ ID NOS: 50231  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 414  
LENGTH: 6274  
TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(6274)  
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-3)  
US-10-767-471-414

Query Match 10.4%; Score 151.6; DB 6; Length 6274;  
Best Local Similarity 52.4%; Pred. No. 5.2e-31;  
Matches 446; Conservative 1; Mismatches 371; Indels 33; Gaps 4;

20 TGCCCAACGACTTGAATGGGCTTCGCAACGCGCCGCTACAGATGGAAGCGCCCTCA 79  
4141 TTCTGAGGCTTCACTGTGAGAGCAAGCTTCTGCTGATATCAGATGGAAGTGGCGTGA 4200  
80 AAGAAGTGGCCGCGCCCTCCATCTGGGACAGTACTGCACTGCACTGGAGCCATCGCGCA 139  
4201 GAGCAGATGGCAAGAGCTCAGATTTGGGACAGTTTCTCACACCACTAGAGGTTG 4260  
140 CCAAGCGCCCAACGGGATGTGCTTGGCATCTACCAACGCTACGATGAGGACTTTG 199  
4261 AGAAGATGCTATTTGAGAGAGTGGCTGTCAGATTATCAAGATTGCTGAGGATTTG 4320  
200 ATCTTGAACCAAGTACGCGGCAAGGCTTACCGCTTCTCTGCTGGTGGCGATCA 259  
4321 TCACCTGCAAGACTGGGAGTGTCCCACTACCGTTTTCATCTCTGCTGCTGATCC 4380  
260 TTCCCTCGGCGGACGCTGATCCGTCACAGAGGGAATTGATTTTACAGCAAC 319  
4381 TCCCTGATGACACCAAGT---CATCATTAACGCGGCTGAACTACTAGTGGAGC 4437  
320 TGAATGACGCTTGTAGGCGGGGATTCAGGCTTGGGATCTTTGATCAGCTGGATC 379  
4438 TCAATGATACCTGCTGCGCGCAAGCTTCAAGCTTCAAGTATTTACATCTGGAGC 4497  
380 TGCTCAGGCGCTTACGATCGCTATGAGGCTGGCTCAACGTCGGAAGAGTCCAGCTGG 439  
4498 TACACAGAGCGTCCAAAGT---GTAGAGGCTGGGAGATGAGAACCATGTCGAGCGG- 4553  
440 ACTTGAAGCGTATGCGAGGTTGCTTTGAAAGTTTGGGAGACGAGTCAGAACTGA 499  
4554 -TTTAAAGGATATGAGATGTCTTCCAGAGGCTGGGAGCAAGGTAAATTTTGA 4611  
500 TCACCATCAACGAGCCCTGATTCAGGCTATGATGATGTCACCGGAGCAACGCC 559  
4612 TCAGTGAATGAGCCCTTTGATTTGATTCAGAGGCTATGCTACGGAACAGAGCTC 4671  
560 CGGCGAGGACAGCATTAACAGACATCCACCGAGGCAACATGCACTGAGCGCTGGC 619  
4672 CAGGAGTCTCA-----ATAGGCTGGGACATGCGCCCTACA 4707  
620 TCGCTGAAAGGCCAGATGATGAGCCATGCGCGCGCTGCTACAGAGGACT 679  
4708 TTGTGGCCCAATTAATTAAGGCTCATGCTAGGCTGGCAATCTGTACAAAGATGTG 4767  
680 TTGGCCCTGCAAGAGGCGCAGATGCGCATCTGCTCAACGCGGACTATGAGGCTT 729  
4768 ACCGCGCACTCAAGGTGGCTGATTTTCAATCACTACAGATGATGGGCTGAAACCA 4827  
740 GGAACAGATGAGCTTCGGAACAGAGGCTGTCAGGACAGATGAAATTTTCAATTG 799  
4828 GAGATCCCTCTAACAAGAGAGATGTGAGGACACAGAGATATGTTGATGAGAG 4887  
800 GCTGCTTGGCAATTCCTATTTTGAAGAGACTATTCAGAGAGACTGAAGAGAGC 859  
4888 GCTGCTTGGCAATTCCTATTTTGAAGAGACTATTCAGAGAGACTGAAGAGAGC 4947  
860 TGGGCGAGAG 870  
4948 TCCGTGACAG 4958

RESULT 5

```
US-10-767-701-6851
; Sequence 6851, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6851
; LENGTH: 767
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; NAME/KEY: unsure
; LOCATION: (1)..(767)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS5746_1
US-10-767-701-6851
```

```
Query Match      10.2%; Score 147.8; DB 6; Length 767;
Best Local Similarity 55.4%; Pred. No. 2.5e-30;
Matches 352; Conservative 0; Mismatches 269; Indels 14; Gaps 3;
```

```
QY 16 GCTGCGCCCAAGCATTTGAAATGAGGGCTTCGCAAGCGCCCTCAACATGGAAGCGCC 75
DB 147 GCGTCCCGGAGGGCTTGCTTTCGGAGACGGCGGTGCGTACAGGTGAGGGGATG 206
QY 76 GTCAAGAAGGTGCGCGCGCCGCTCATCTGGAGACAGTACTGCGCACCTGAGGCATCG 135
DB 207 GCCAGACAGCGGGGCGCGGGGCCAGCATCTGGAGACGCTTCATAGAGTACCGCGGACC 266
QY 136 CGCAGCAACGGGCGCCAGCGGATGTGGCTTGCATCTACACCGCTGATGAGAGAC 195
DB 267 ATCCCTAACAATGCGACCGCTGATGTGACGGTGCAGATGATCACTGGTACAGGAGAT 326
QY 196 TTTGATCTCTTGACCAAGTACGGCGCAAGGCTTCTCTGTCGTGGTGGCGG 255
DB 327 GTGAACATATGAAAGAAATGAGGCTTTGATGCGTACCGGTTTCAATTTCTTGTCGAGG 386
QY 256 ATCATTTCCCTCGCGGAGGCTGATCCGTCACAAGAGAGGAAATGATTTTACAGC 315
DB 387 ATTTCCCAATGGAATGCAAG-----GTGAACCAAGAGAGGATGATTTACTATAC 440
QY 316 AACATGATGACGCGCTGTTGAGCGGGGATTCAGCGCTTGGGTGACTTTTACCACTGG 375
DB 441 AGGCTCATATGATTCATGCTTCAGAAAGTATCAAGCGCGTATGCAAAATCTTACCATTTAT 500
QY 376 GATCTGCTCAAGGCGCTTCAGATGCTATGAGAGGCTGCTCAACGTGGAAGAGTCCAG 435
DB 501 GACCTCCCATTTGGGCACTCTCATGAAACAGTACTTGGGCTGGCTTACG---CCAAAGATTGTG 557
QY 436 CTGACTTTGAGCGGATGCAAGGTTGTGCTTTGAACGTTTGGGAGCCGATCCAGAAC 495
DB 558 GAGGCGTTTGAAGATGACGCGGAGTTCGCTTCCAGAGCTTCGAGACAGAGGTGAAGAC 617
QY 496 TGAATCAACATCAAGMCCCTGATTCAGGCGCATATGATATATGCAACCGGAGACAC 555
DB 618 TGGTTTATCTTCAAGAGCGAGGTGCGTCTCTTGGCTAAGCAATGGCTTACAC 677
QY 556 GCCCGGAGAGAGCAGCATTTACAGACACTCCAGGAGGACACATGCGCATGAGCGG 615
DB 678 GCANCCGG-----GAAAGTGTTCGAGTGGCGCTCCGAGGACATCTCAGACGAGACCG 732
QY 616 TGGCTCGTGAAGAGGCCAGATCATGACCATTC 650
DB 733 TACCTGCTGACACATCTCATCTTTCATATGC 767
```

```
RESULT 6
PCT-US04-02242-93
; Sequence 93, Application PC/TUS0402242
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
; TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANI
; FILE REFERENCE: 56446-20118.40
; CURRENT APPLICATION NUMBER: PCT/US04/02242
; PRIOR FILING DATE: 2004-01-26
; PRIOR APPLICATION NUMBER: 60/442,794
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 93
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-93
```

```
Query Match      9.6%; Score 139; DB 1; Length 1389;
Best Local Similarity 57.1%; Pred. No. 7.2e-28;
Matches 316; Conservative 0; Mismatches 225; Indels 12; Gaps 3;
```

```
QY 11 CGTAGCTCTGCCCAAGCATTTGAAATGAGGGCTTCGCAAGCGCCCTTACAGATGGAAG 70
DB 26 CCGTCTCTTCCAGAGCAATTCGTCTGGGTCTCTCGCGGCTCTTCAAGTGAAG 85
QY 71 GCGCGCTCAAGAAGGTGCGCGCGCCGCTCATCTGGGACAGATCTGCACTGAGAG 130
DB 86 GCGCGCTCAAGAGAGAGGAGGAGGCGCTCTCTGAGACATTTCTGGAAGACCGG 145
QY 131 CATGCGCACCAACGCGCGCCAGCGGATGTGGCTTGCATCTACCAACCGCTACAGATG 190
DB 146 GAGCGGCTTCCAGAGGAGCAGACGCGGCGGTGCTTGCACACTATACCGCTACGAG 205
QY 191 AGACATTTGATCTCTTGACCAAGTACGGCGCAAGGCTTACCGCTTCTCTTGTGCTGT 250
DB 206 AGACGCGGTGATGACAGAGTGGCTGACCGCTTACCGCTGAGCGTGTGCTGGC 265
QY 251 CGCGATCATTTCCCTCGCGGCGAGGCTGATCCGTCACAAGAGAGGAAATGATTTT 310
DB 266 CCGAGTGTCTCCGAGAGGGGTGCGGAG-----CCAAAGAGAGGCGCTGACTTCT 319
QY 311 ACAGCAAACTGATTAAGCGCCCTGTTGAGCGGGGATTCAGCGCTTGGGTGACTTTGATCC 370
DB 320 ACTCGCGTTGATGACGCGGCTGCTGAGGAGGATTTACGCGCTTGGTTAACGCTTTTTC 379
QY 371 ACTGGATTCGCTCAGGCGGCTTCAAGATCGTATGAGGCTGAGCTCAACGTGGAAGAG 430
DB 380 ATTGGAGTACCCCTTGGGCTT---CTATACCGGGGGGCTGGCTCAACCGGAGATGCG 436
QY 431 TCCAGCTGACCTTTGAGCGGATGCGAGGTTGTCTTTGAACGTTTGGGAGCCGAGTCC 490
DB 437 CGGATTGG---TTTGGCGAGTACGCGGGCTGATATGCGCATGTGCTCCGACCGGGTGC 493
QY 491 AGAATGATCACCATCAACGAMCCTTGATTCAGGCGCATATGATATATGCAACCGGCA 550
DB 494 AGCATTTCTTCACTCAAGAGAGGCCAGGCTTATATCGCTTGGACACCTCGAGGGTA 553
QY 551 GCAAGCGCCCGG 563
DB 554 AGCATCTCAGG 566
```

```
RESULT 7
US-10-767-701-14696
; Sequence 14696, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```

APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 14696  
LENGTH: 2042  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS4596\_1  
US-10-767-701-14696

Query Match 9.0%; Score 131; DB 6; Length 2042;  
Best Local Similarity 51.5%; Pred. No. 1,2e-25;  
Matches 433; Conservative 1; Mismatches 386; Indels 21; Gaps 5;

20 TGGCCACGACTTGTATGGGGCTTGCACAGCGCGCTCCAGATGAGAGGGCGCTCA 79  
165 TCCCGCGGGGCTTCTTGGCGCTGCTGCGCTACAGGTTGAAGTGCACTTG 224  
80 AAGAGGTGGCGCGCGCGCTCCATCTGGACACGTAAGTCCCACTTGAAGCACTCGCA 139  
225 CAGAGATGAG 281  
140 CCAAGCGCGCAAGCGAGATGAGCTTGCATCAACACCGCTACGATGAGAGACTTG 199  
282 TTGACCGTGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
200 ATCTTTCACCAAGTACGCGCGCAAGCGCTTCCCTCTCTCTCTCTCTCTCTCTCT 259  
342 AGCTTTTGATGAGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 401  
260 TTCCTCGCGCGCGCGCGCTGATCCCTCAACGAGAGAGAGAGAGAGAGAGAGAG 319  
402 TTCCTGATGCG-----CGAGAGCTGTCAATCCGAGGGGCTGAGTACTCAACATC 455  
320 TGTATGACGCGCTTGTGAGCGGGGATACAGCGCTTGGTGAATTTTATCACTGGATC 379  
456 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 515  
380 TGGCTCAGGCGCTTCAAGATGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439  
516 TTCCTCAGGCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 572  
440 ACTTTGAGCGTATGAG 499  
573 ACTACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 632  
500 TCACCATCAACGAGCGCTGATTCAGGCTATGATGATGATGATGATGATGATGATG 559  
633 GCACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692  
560 C-----GGGCGAGAGAGAGATTAACAAGACATCCAGAGAGAGAGAGAGAGAG 613  
693 CAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 752  
614 CGTGGCTGCTGAGAAAGAGCGCATGATGATGATGATGATGATGATGATGATGATG 673  
753 CGTACATGAG 812  
674 GGGACCTTTCG 733  
813 AGAAGTACAG 872  
734 AGGCGTGGAG 793  
873 AGGCTT---GCAACGAG 929  
794 ACATTGGCTGTTTGGCAATCCATCTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 853

Db 930 ACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 989  
Qy 854 A 854  
Db 990 A 990

## RESULT 8

US-10-093-037A-2  
Sequence 2, Application US/10093037A

GENERAL INFORMATION:

APPLICANT: Jay M. Short

APPLICANT: Bylina, Edward

APPLICANT: Swanson, Ronald V.

APPLICANT: Mathur, Eric J.

APPLICANT: Lam, David E.

TITLE OF INVENTION: ENZYMES HAVING GLYCOSIDASE ACTIVITY AND METHODS OF USE THEREOF

FILE REFERENCE: 56462001402

CURRENT APPLICATION NUMBER: US/10/093,037A

CURRENT FILING DATE: 2002-03-06

PRIOR APPLICATION NUMBER: US 09/910,579

PRIOR FILING DATE: 2001-07-20

PRIOR APPLICATION NUMBER: US 09/134,078

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 08/949,026

PRIOR FILING DATE: 1997-10-10

PRIOR APPLICATION NUMBER: US 60/056,916

PRIOR FILING DATE: 1996-12-06

NUMBER OF SEQ ID NOS: 72

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 1317

TYPE: DNA

ORGANISM: Thermotoga sp.

US-10-093-037A-2

Query Match 9.0%; Score 130.8; DB 6; Length 1317;

Best Local Similarity 56.4%; Pred. No. 1,1e-25;

Matches 309; Conservative 1; Mismatches 223; Indels 15; Gaps 3;

16 GCTTCGCCAAGACATTTGATGAGGGCTTGCACAGCGCGCTTACAGATGAGAGCGC 75  
16 GATTTTCAAAAGATTTTATCTTGGAGCGCTTACAGAGATGAGATGAGATGAGATG 75  
76 GTCAAG 135  
76 GCAAG 135  
136 CGACACAG 195  
136 ACCCTGACGCTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 195  
196 TTTGATCTCTTGAACAAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255  
196 ATCCAGTATGAG 255  
256 ATGATTCCTCG 315  
256 ATTATGCGAGATG-----GAAAGACATCAACCAAGAGAGAGAGAGAGAGAG 306  
316 AAATGATGAG 375  
307 AGACTGCTGATGAG 366  
376 GATTCGCTCAG 435  
367 GACTTACCTTACAG 420  
436 CTGAGCTTGAAG 495  
421 CTCTATTTAG 480



Db 34159 GGCAGAGCAGGAAGGGGGTCTCGCTGAGCCCTCAGTACACACTGGGAGAGCCCAAGT 34218  
Qy 743 ACAGCATATGAGCTCGGACACAAAGAGAGGCTGTGAGCAGCGATGGAATTTTCAATTGGCT 802  
Db 34219 CACCAAGGGTCCCAAGAGATGTGAGAGCCCTGACCGAATGTGCACTTCTCCCTGGGCT 34278  
Qy 803 GATTGCGCAATCCCATCTTCTTGAAGAAGACTATCCAGAGAGCATGAGAGAGCAGCTGG 862  
Db 34279 GATTGCTACCCCATTTTGAAGAAGCACTATCTGACACATGAAAGTGAAGTGG 34338  
Qy 863 GCGAGAGG 870  
Db 34339 GGAACAGG 34346

RESULT 11  
US-10-767-701-8210  
; Sequence 8210, Application US/10767701  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 8210  
; LENGTH: 689  
; TYPE: DNA  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS33734\_1  
; US-10-767-701-8210

Query Match 8.6%; Score 125.2; DB 6; Length 689;  
Best Local Similarity 56.4%; Pred. No. 2,9e-24;  
Matches 299; Conservative 0; Mismatches 218; Indels 12; Gaps 3;  
Qy 12 GCTAAGCTTCCCAACGACTTTGAATGGGCTTTCGCAAGCGCCCTACCAATCGAAG 71  
Db 173 GCAGGATTTTCCAGGGCTTTCGTTGGGCGCGGACATCAGCTTATCAGTACGAAG 232  
Qy 72 CGCCGCTCAAGAGAGTGGCCGCGCCGCTCATCTGGGACACGTAAGTCCACCTGGAGCC 131  
Db 233 GCGAACCGATGAGATGGAAGAGGAGCCCAAGCATATGGACACGTTTACTATGCAAGGAG 292  
Qy 132 ATCGCCGACCAACGCGCCCAAGCGGATGTGGCTTGGATCACTACCAAGCGCTACATGA 191  
Db 293 GATGCCGACAAAGACT---GGCATCTAGGGGAGAGCGGCTTACCAAAATACAGGA 349  
Qy 192 GGAATTGATCTTGTGACCAAGTACGCGCAAGGCGCTTACCGCTTCTCTTGTGCGTGC 251  
Db 350 AGATGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 409  
Qy 252 GCGGATCATTCCTCGCGCGGAGGCTGATCCCGTCAAGAGAGGAATTTGATTTTA 311  
Db 410 CAGGCTTCTTCAAGAGGAG-----AGGACCATCAACCCCAAGGCTTTCAGATTTA 463  
Qy 312 CAGCAAACTGATGACGCGCTTGTGAGGCGGGGATCAGCGCTTGGGCTTGTGTGCA 371  
Db 464 CAGCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 523  
Qy 372 CTGGAATCTGCTCAGCGCTTCAAGATCGCTATGAGAGCTGTGCTCAACGTGAAGAGT 431  
Db 524 CTGGAATCTGCTCAGATCTGAGAGGAGTACATGAGCTGTGCTGAGCCCGAGGAGT 583  
Qy 432 CAGGCTGATTTGAGCGGATGAGAGGTTGCTTGAACGTTTGGGAGCGAGTCCA 491  
Db 584 ---GAGAGCTTCAAGCGGCTACGCGAGCGTGTCTTCCGGAGTTCGCGACCGGGTGA 640

Qy 492 GAATGATCACCATCAAGAGCCCTGATTTAGGCCCATCTATGATA 539  
Db 641 GCACTGAGACCATGAGAGCGCAAGCATGATCTCATCGCGGCTTA 688

RESULT 12  
US-10-767-795-4264  
; Sequence 4264, Application US/10767795  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(53534)B  
; CURRENT APPLICATION NUMBER: US/10/767,795  
; CURRENT FILING DATE: 2004-01-30  
; NUMBER OF SEQ ID NOS: 117596  
; SEQ ID NO 4264  
; LENGTH: 649  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C301\_1  
; US-10-767-795-4264

Query Match 8.2%; Score 119.4; DB 6; Length 649;  
Best Local Similarity 54.5%; Pred. No. 1e-22;  
Matches 285; Conservative 1; Mismatches 227; Indels 10; Gaps 2;

Qy 18 TCTGCCCAACGACTTTGAATGGGCTTTCGCAAGCGCCCTACAGATCGAGCGCCGT 77  
Db 137 TTTCCGAGGAGATTTGTGTGGAATCGCACTTCACTTATCAAGTGAAGATGGC 196  
Qy 78 CAAGAAGGTGGCCGCGCCCTCCATCTGGGACACGTAAGTCCACCTGAGCCATCGC 137  
Db 197 TAATAAAGGCGCGGAGACCTTGCAATTTGGGATGTTTATGTTAAACAACGACCATAT 256  
Qy 138 CACCAACGCGGACCAAGGGATGTGGCTTGGCATCACTACCGGTACGATGAGACTT 197  
Db 257 TGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 316  
Qy 198 TGATCTCTTGAACCAAGTACGCGCAAGGCTTCTCTTGTGCTGCTGCGGAT 257  
Db 317 AGATTTGTGGCAAAATTTAAATTTGATCTTATGCTTCTCAATTCATGATGCAAGAT 376  
Qy 258 CATTCCTCGCGGAGGCTGATCCCGTCAAGAGAGGAATGAGTTTATACAGCA 317  
Db 377 CTTTCCAGAGGAGT-----TTGAAAGATTAATTGGAGGAGTGAATTATTAACAAG 430  
Qy 318 ACTGATGACGCGCTTGTGAGGCGGGTATCAAGCGCTTGGTGAATTTGATACACTGGGA 377  
Db 431 GTTGATCAATTAATTTGTTAAGAAAGTATATCTCGTATGGAACCTTGATACATTATGA 490  
Qy 378 TCTGCTCAGGCGCTTACAGATGCTATGAGGCTGCTCAACGTGGAAGAGTCCAGCT 437  
Db 491 TCTCCTCAAGCTTCTCAAGAGATGATGAGTTCCTTGAACCGTCAAA---TTGTGGA 546  
Qy 438 GGAATTGAGCGGATGAGAGTGTGCTTGAACGTTTGGGAGCCAGTCCAGAACTG 497  
Db 547 AAATTAATGCTGATTAAGCAAGATTTGTTTCAAAAATTTGGGGGATTAAGTCAAAAAC 606  
Qy 498 GATCACCATCAAGAGCCCTGATTAAGGCCATCTATGATAT 540  
Db 607 GTTACATTAATGAACCAAGATCGGGGTGTTTGTGTTT 649

RESULT 13  
US-10-767-701-12773  
; Sequence 12773, Application US/10767701  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 12773  
LENGTH: 716  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS25405\_1  
US-10-767-701-12773

Query Match 7.9%; Score 115.4; DB 6; Length 716;  
Best Local Similarity 56.6%; Pred. No. 1.3e-21;  
Matches 237; Conservative 0; Mismatches 176; Indels 6; Gaps 1;

QY 20 TGCCCAAGCACTTGAATGGGCTTCGCAACGGCGCCCTACCAAGATCGAAGGCGCGTCA 79  
DB 281 TCCCAAGGGGTTGTTGTTGGGACGGGACGTCGGCTACCAAGTCGAGGCGCCCGCT 340  
QY 80 AAGAAGGTGGCGCGCCCGCTCATCTGGGACACGTACTGCCACCTGGAGCCATCGCGCA 139  
DB 341 CCAACCAAGCGCGGAGACCTCCATCTGGGATTCATTCGCGCACGTCGCCAAGAAATATG 400  
QY 140 CCAAGCGGCCAAGGGGATGGCTTGCATCTACCTACCCGCTACAGTAGAGACTTTG 199  
DB 401 CGGGAAATCAAAATGGAGACCTTCAGTAGATCAATCCATCGCTACCAAGAAATATG 460  
QY 200 ATCTCTGACCAAGTACGAGCGCAAGGCTACCGCTTCTCTTCTGTTGTCGGGATCA 259  
DB 461 ATCTCATGAGAGATTGAATTTGATGCTACCGGTTTTCATCTCGTGTCGAGATCT 520  
QY 260 TTCCCTCGGCGGACGCTGATCCCTCAACGAGAGGGAATTGATTTTACGCAAC 319  
DB 521 TCCAGATGGCG-----AGGGAAAGTCATCCAGAAAGGTGATATTAACAATATC 574  
QY 320 TGATGACGCGCTTGTGAGGGGGTATCACGCTTGGGTGACTTTGATCCAGCGGATC 379  
DB 575 TGATTAACATATCTCTTCAGAAAGCATGACTCTTACATACCTTTACCATATGATC 634  
QY 380 TGCCTCAGGCGCTTACGATGCTATGAGGCTGCTCAACGTCGAAAGGATCCAGCTG 438  
DB 635 TTCCCTTGGCTTGAGAGAAATATGAGGCTTGAAGCGCTAAAGATGCGGACCTG 693

RESULT 14  
US-10-767-701-26695  
Sequence 26695, Application US/10767701  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5353)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 26695  
LENGTH: 637  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: 30977777  
US-10-767-701-26695

Query Match 7.9%; Score 114.8; DB 6; Length 637;  
Best Local Similarity 58.4%; Pred. No. 1.7e-21;  
Matches 222; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 16 GCTCTGCCCAACGACTTTGAATGGGCTTGCACACGGCGCCCTACCAAGATCGAAGCGCC 75  
DB 256 GCGTTCCCAAGGGGTTTCACTTCGGGACGGCCAGCGTCGCTTCCAGTTCGAGGGCGG 315  
QY 76 GTCAAGAGGTGGCGCGCCCGCTCATCTGGGACACGTAATGCACTCGAGCCATCG 135  
DB 316 GCGACGTCCGCGCGCGCGCGCCAGCATCTGGATCTCTTCTGTCACACCCCGGAAG 375  
QY 136 CGCACCAAGCGCGCAACGGCGATGGCTTGGCATCTACCAACGCTACGATGAGAC 195  
DB 376 ATGCTGAAGACGGGAACGAGATGTTACACAGATGAATATCATGCTTACAGGAAGAT 435  
QY 196 TTGATCTCTTGAACCAATGACGGCGCAAGAGCTTACCGCTTCTGTCGTGCTCGG 255  
DB 436 GTTATCTCATGAAGAAAGCTTAATTTGATGACATACCGGTTTCAATCTCTGTCAGG 495  
QY 256 ATCATTTCCCTCGGCGGACGCTGATCCCGTCAACGAGAGGAATTGATTACAGC 315  
DB 496 ATCTTCCAGATGGCGGAAGG-----AAGTTAATGAAGAGATACGATATTACAC 549  
QY 316 AAATGATGACGCGCTTGTGAGGCGGATATCACGCTTGGTGACTTTGATCACTGG 375  
DB 550 AATCTTATAGCATATGATTAAGCAAGTCTTACTCTTAACGCAACCTTAACACTAT 609  
QY 376 GATCTGCTCAGGCGCTTCA 395  
DB 610 GATCTTCACTTGGCGCTTCA 629

RESULT 15  
US-10-767-795-4474  
Sequence 4474, Application US/10767795  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53534)B  
CURRENT APPLICATION NUMBER: US/10/767,795  
CURRENT FILING DATE: 2004-01-30  
NUMBER OF SEQ ID NOS: 117596  
SEQ ID NO 4474  
LENGTH: 558  
TYPE: DNA  
ORGANISM: Gossypium hirsutum  
FEATURE:  
OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C30142\_1  
US-10-767-795-4474

Query Match 7.5%; Score 109; DB 6; Length 558;  
Best Local Similarity 58.7%; Pred. No. 6e-20;  
Matches 225; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 16 GCTCTGCCCAACGACTTTGAATGGGCTTGCACACGGCGCCCTACCAAGATCGAAGCGCC 75  
DB 107 GATTTCCCTCCCGCACTTGTCTTCGTTCTTACCTTCTTCAAGATTTGAAGGAGT 166  
QY 76 GTCAAGAGGTGGCGCGCGCGCTCATCTGGGACACGTAATGCACTCGAGCCATCG 135  
DB 167 GTGAACCAAGGTGGCAAGGTAAAGCATATGGATCTTCTTCAATTTGAAGGAAAA 226  
QY 136 CGCACCAAGCGCGCAACGGCGATGGCTTGGCATCTACCAACGCTACGATGAGAC 195  
DB 227 ATGCTTATGAGCAATGCTGATGTTGAGTGAATATTAACAAGATCAAGGAAGAT 286  
QY 196 TTGATCTCTTGAACCAATGACGGCGCAAGAGCTTCTTCTTGTGCTGCTCGG 255  
DB 287 ATAGAGCTTATATCGAGTTAGGTTCAAGCTTACAGATTTTCCATATCATGCTCGT 346  
QY 256 ATCATTTCCCTCGGCGGACGCTGATCCCGTCAACGAGAGGAATTGATTACAGC 315  
DB 347 ATTTTCCAGATG--GTTTGGAAAAACAAAGTTATAGAGGAAGATTGCAATTTACAC 404

Qy	316	AAACTGATTGACCGCCCTGTGAGGGGGGTATCAGCCCTTGGGTGACTTTGTACCACTGG	375
Db	405	AATGTGATCGATGCGCCT-TTTGAAAGGGTATTGAGCCCTTGTGACATTATACCAATTGG	463
Qy	376	GATCTGCTTCAGGGCGCTTCACGA	398
Db	464	GATTCTTCTTGCATCTCCATGA	486

Search completed: March 29, 2004, 02:56:08  
Job time : 126.98 secs

**This Page Blank (uspto)**